

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 18, 2004, 22:37:26 ; Search time 56.0692 Seconds  
(without alignments)  
11771.718 Million cell updates/sec

Title: US-09-729-264-3

Perfect score: 388

Sequence: 1 agtgatcatgtgagcaggag.....gtaatacaactgtagtatag 1168

Scoring table:  
OLIGO Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2094264

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Listing first 135 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2.1/USPTO spoal\_p/US09729264/runat.17092004.155106.2337/app\_query.fasta\_1.4117  
-DB=A\_Geneseq\_29Jan04 -QMT=fastan -SUFFIX=oli.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=135 -DOCALIGN=200 -THR SCORE=quality -THR MIN=0 -ALIGN=45 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09729264 @CGN 1.1 321 @runat.17092004.155106.2337 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAD=1 -XGAPOP=60 -XGAEXT=60 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=60 -YGAEXT=60 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	386	99.5	386	5 AAU75541	Aau75541 Human B7-
2	257	66.2	382	5 AAU75540	Aau75540 Human B7-
3	199	51.3	386	5 AAU75542	Aau75542 Human B7-
4	199	51.3	407	7 ADB64920	Adb64920 Human pro
5	142	36.6	377	5 AAU75543	Aau75543 Human B7-
6	142	36.6	463	4 ABC28169	Abc28169 Novel hum
7	12	3.2	78	4 AAM20679	Aam20679 Peptide #
8	12	3.2	78	4 ABB42042	Abb42042 Peptide #
9	12	3.2	78	4 AAM35844	Aam35844 Peptide #
10	12	3.2	78	4 ABB25654	Abb25654 Protein #

C 11	12	3.2	78	4 AAM75735	Aam75735 Human bon
C 12	12	3.2	78	4 AAM62923	Aame2923 Human bra
C 13	12	3.2	78	4 ABG57473	Abg57473 Human liv
C 14	12	3.2	78	5 ABG45219	Abg45219 Human pep
C 15	11	2.8	524	4 AAU07370	Aau07370 G protein
C 16	11	2.8	1805	4 ABB65262	Abb65262 Drosophil
C 17	10	2.6	85	3 AAY65872	Aay65872 n-myc mut
C 18	10	2.6	94	3 AAY65874	Aay65874 n-myc mut
C 19	10	2.6	95	3 AAY65875	Aay65875 n-myc mut
C 20	10	2.6	2168	4 ABB64563	Abb64563 Drosophil
C 21	9	2.4	96	4 ABB69039	Abb69039 Drosophil
C 22	9	2.4	191	4 ABB63572	Abb63572 Drosophil
C 23	9	2.4	191	4 ABB70426	Abb70426 Drosophil
C 24	9	2.4	191	4 ABB67144	Abb67144 Drosophil
C 25	9	2.3	253	6 ADA48420	Ada48420 Rice prot
C 26	9	2.3	258	6 ABUI7237	Abui7237 Protein e
C 27	9	2.3	269	6 ADA36602	Ada36602 Acinetoba
C 28	9	2.3	331	2 AAR22653	Aar22653 Transglut
C 29	9	2.3	331	2 AAY33665	Aay33665 Streptove
C 30	9	2.3	331	2 AAY33662	Aay33662 Streptove
C 31	9	2.3	331	2 AAM67770	Aam67770 A transgl
C 32	9	2.3	331	4 AAB81161	Aab81161 Transglut
C 33	9	2.3	331	5 ABB06742	Abb06742 Streptove
C 34	9	2.3	332	2 AAM67771	Aaw67771 Protein e
C 35	9	2.3	332	3 AAB12809	Aab12809 Transglut
C 36	9	2.3	332	5 ABB06743	Abb06743 Streptove
C 37	9	2.3	332	7 ADB37643	Adb37643 Streptomy
C 38	9	2.3	346	2 AAR49048	Aar49048 Bacterial
C 39	9	2.3	355	7 ADB37647	Adb37647 Streptomy
C 40	9	2.3	360	4 ABG22212	Abg22212 Novel hum
C 41	9	2.3	406	2 AAR22651	Aar22651 Transglut
C 42	9	2.3	407	4 AAB97831	Aab97831 S. mobara
C 43	9	2.3	407	4 AAB81166	Aab81166 Prepro-tr
C 44	9	2.3	407	6 ABU07390	Abu07390 Foreign p
C 45	9	2.3	408	2 AAR49049	Aar49049 RTG-contg
C 46	9	2.4	494	4 ABB68018	Abb68018 Drosophil
C 47	9	2.4	567	4 ABB65901	Abb65901 Drosophil
C 48	9	2.4	567	4 ABB64078	Abb64078 Drosophil
C 49	9	2.4	567	4 ABB66007	Abb66007 Drosophil
C 50	9	2.3	820	7 ADE65531	Ade65531 Streptove
C 51	9	2.4	904	4 ABB71477	Abb71477 Drosophil
C 52	9	2.4	904	5 ABG70014	Abg70014 Larval vi
C 53	9	2.4	908	6 ABU08492	Abu08492 Alpha-hel
C 54	9	2.4	1076	6 ABU81145	Abu81145 Human PRO
C 55	9	2.4	1076	6 ABU66845	Abu66845 Human PRO
C 56	9	2.4	1301	4 ABB65129	Abb65129 Drosophil
C 57	9	2.4	1379	4 ABB68940	Abb68940 Drosophil
C 58	9	2.4	1601	4 ABB60248	Abb60248 Drosophil
C 59	9	2.4	3972	3 AAB23749	Aab23749 S. avermi
C 60	9	2.4	3972	4 AAG65264	Aag65264 Streptomy
C 61	9	2.4	3972	4 AAG65268	Aag65268 Streptomy
C 62	9	2.4	4440	6 ABU88256	Abu88256 Novel hum
C 63	9	2.4	4440	6 ABU90135	Abu90135 Novel hum
C 64	9	2.4	4440	6 ABU96437	Abu96437 Novel hum
C 65	9	2.4	4440	6 ABU99046	Abu99046 Novel hum
C 66	9	2.4	4440	6 ABU98261	Abu98261 Novel hum
C 67	9	2.4	4440	6 ABU91967	Abu91967 Novel hum
C 68	9	2.4	4440	6 ABU85271	Abu85271 Novel hum
C 69	9	2.4	4440	6 ABO00410	Ab000410 Novel hum
C 70	9	2.4	4440	6 ABU88961	Abu88961 Novel hum
C 71	9	2.4	4440	6 ABO06457	Ab006457 Novel hum
C 72	9	2.4	4440	6 ABU95517	Abu95517 Novel hum
C 73	9	2.4	4440	6 ABU95207	Abu95207 Novel hum
C 74	9	2.4	4440	6 ABU90755	Abu90755 Novel hum
C 75	9	2.4	4440	6 ABU93917	Abu93917 Novel hum
C 76	9	2.4	4440	6 ABU86191	Abu86191 Novel hum
C 77	9	2.4	4440	6 ABU82046	Abu82046 Novel hum
C 78	9	2.4	4440	6 ABU07907	Abu07907 Novel hum
C 79	9	2.4	4440	6 ABU94227	Abu94227 Novel hum
C 80	9	2.4	4440	6 ABO00100	Ab000100 Novel hum
C 81	9	2.4	4440	6 ABU87111	Abu87111 Novel hum
C 82	9	2.4	4440	6 ABU91352	Abu91352 Novel hum
C 83	9	2.4	4440	6 ABU90445	Abu90445 Novel hum

C 84 9 2.4 4440 6 ABU97036 Novel hum  
C 85 9 2.4 4440 6 ABO05232 Novel hum  
C 86 9 2.4 4440 6 ABO23752 S. avermi  
C 87 9 2.4 5532 3 AAG65267 Streptomy  
C 88 9 2.4 6025 4 AAG84939 Shrimp wh  
C 89 9 2.4 6077 6 ABP95996 White spo  
C 90 8 2.1 36 4 ABB15591 Human ner  
C 91 8 2.1 48 2 AAY12003 Human 5'  
C 92 8 2.1 50 4 AAM21199 Peptide #  
C 93 8 2.1 50 4 ABB43522 Peptide #  
C 94 8 2.1 50 4 AAM37416 Peptide #  
C 95 8 2.1 50 4 ABB26481 Protein #  
C 96 8 2.1 50 4 AAM77263 Human bon  
C 97 8 2.1 50 4 AAM64459 Human liv  
C 98 8 2.1 50 4 AAG58900 Human liv  
C 99 8 2.1 50 5 AAG46280 Human pep  
C 100 8 2.1 79 5 ABP05686 Human ORF  
C 101 8 2.1 81 4 ABB70231 Drosophil  
C 102 8 2.1 82 3 AAY76046 Murine sk  
C 103 8 2.1 82 4 AAB55985 Skin cell  
C 104 8 2.1 82 5 ABB72185 Murine pr  
C 105 8 2.1 89 4 AAM18746 Peptide #  
C 106 8 2.1 89 4 ABB37844 Peptide #  
C 107 8 2.1 89 4 AAM31246 Peptide #  
C 108 8 2.1 89 4 ABB23106 Protein #  
C 109 8 2.1 89 4 AAM70960 Human bon  
C 110 8 2.1 89 4 AAM58466 Human bra  
C 111 8 2.1 89 4 AAG52681 Human liv  
C 112 8 2.1 89 5 ABB40758 Human pep  
C 113 8 2.1 89 5 AABF32378 Human red  
C 114 8 2.1 100 3 AAB40618 Human ORF  
C 115 8 2.1 100 5 AABP02711 Human ORF  
C 116 8 2.1 106 3 AAB42377 Human ORF  
C 117 8 2.1 119 4 ABB69357 Drosophil  
C 118 8 2.1 134 4 ABB63906 Drosophil  
C 119 8 2.1 140 4 ABB71305 Drosophil  
C 120 8 2.1 148 5 AAU96222 Human sec  
C 121 8 2.1 148 5 ABB64793 Human alb  
C 122 8 2.1 149 4 ABB69455 Drosophil  
C 123 8 2.1 150 4 AAU43063 Propionib  
C 124 8 2.1 150 5 ABB61831 Propionate  
C 125 8 2.1 150 6 ABBM39582 Propionib  
C 126 8 2.1 153 7 ADC08113 Rice prot  
C 127 8 2.1 155 4 AAB68931 Neisseria  
C 128 8 2.1 165 4 ABG29283 Novel hum  
C 129 8 2.1 181 4 ABG28127 Novel hum  
C 130 8 2.1 184 4 ABB66043 Drosophil  
C 131 8 2.1 185 4 AAB95364 Human pro  
C 132 8 2.1 220 4 AAU51099 Propionib  
C 133 8 2.1 220 6 ABM47618 Propionib  
C 134 8 2.1 223 5 AAU75546 Mouse B7-  
C 135 8 2.1 224 6 ADA48140 Rice prot

## ALIGNMENTS

## RESULT 1

AAU75541  
ID AAU75541 standard; protein; 386 AA.

XX AC AAU75541;

XX DT 23-APR-2002 (first entry)

XX DE Human B7-like protein, B7-L\_h2.

XX Human; B7-like protein; B7-L; antinfertility; gynaecological;  
KW antitumour; cytostatic; immunosuppressive; antiarthritic; antineuritic;  
KW antinflammatory; dermatological; antipsoriatic; neuroprotective;  
KW antidiabetic; haemostatic; antithyroid; antituber; antiallergic;  
KW antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;  
KW reproductive disorder; graft versus host disease; autoimmune disease;  
KW toxic shock syndrome; allergy; nephropathy; skin disorder;

KW endocrinopathy; lymphoproliferative disorder.

XX Homo sapiens.

PN WO200200710-A2.

XX 03-JAN-2002.

XX 28-JUN-2001; 2001WO-US020719.

PR 28-JUN-2000; 2000US-0214512P.

PR 28-NOV-2000; 2000US-00723264.

(AMGE-) AMGEN INC.

XX Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;

XX WPI; 2002-130881/17.

XX N-PSDB; ABK13029.

XX New B7-like polypeptides, polynucleotides and their modulators, useful  
for diagnosing, preventing and treating reproductive, immune and  
proliferative disorders, e.g. cancer and arteriosclerosis.

XX Claim 13; Fig 2; 135pp; English.

XX The invention relates to an isolated B7-like (B7-L) polypeptide (I). The  
polypeptide, polynucleotide encoding it and antibody against (I) are  
useful for treating B7-like polypeptide-related disease, disorders or  
conditions including reproductive disorders (e.g. infertility,  
miscarriage, preterm labour and delivery and endometriosis) and  
proliferative disorders. Antibodies, soluble proteins comprising  
extracellular domains and other regulators of B7-L polypeptides are  
useful for enhancing the immune response to tumours. (I) plays a role in  
growth and maintenance of cancer cells based on the observation of  
seminal vesicle hyperplasia in transgenic mice overexpressing B7-L  
polypeptide. Hence modulators of (I) are useful for the treatment of  
cancer including seminal vesicle cancer, lung, brain, breast, ovarian,  
testicular cancer and cancers of haematopoietic system. B7-L polypeptide  
pathway can be manipulated to regulate cytotoxic T-lymphocyte response in  
allograft transplantation, graft versus host disease, T-cell dependent B-  
cell mediated diseases and autoimmune diseases. B7-L molecules are useful  
for alleviating the symptoms associated with diseases involving chronic  
immune cell dysfunction or to treat autoimmune diseases such as systemic  
lupus erythematosus, rheumatoid arthritis, multiple sclerosis, diabetes,  
immune thrombocytopenic purpura and psoriasis, chronic inflammatory  
disease such as inflammatory bowel disease (Crohn's disease) and  
diabetes mellitus. They are also useful as immunosuppressive agents for  
bone marrow and organ transplantation or to prolong graft survival. B7-L  
molecules are also useful for diagnosis and treatment of diseases  
involving abnormal cell proliferation, including arteriosclerosis and  
vascular restenosis. Antagonists of B7-L polypeptides are useful for  
alleviation of toxic shock syndrome or allosensitisation due to blood  
transfusions, and for treatment of allergy, asthma and hypersensitivity  
reactions, nephropathies (e.g. glomerulonephritis), skin disorders  
(pemphigus and pemphigoid), endocrinopathies (Grave's disease), various  
pneumopathies (extrinsic alveolitis), vasculopathies, coeliac disease,  
anaemia, thrombocytopenias, Guillain-Barre syndrome and myasthenia  
gravis, and lymphoproliferative disorders such as multiple myeloma. The  
present sequence represents the amino acid sequence of human B7-L\_h2

XX Sequence 386 AA;

XX Alignment Scores:

Pred. No.:	0	Length:	386
Score:	386.00	Matches:	386
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.48%	Indels:	0
DB:	5	Gaps:	0

US-09-729-264-3 (1-1168) x AAU75541 (1-386)

QY 8 ATGGTGGCAGGAGCCATGGAATAATAGAGACCCACCGGTTCTGGTCTCTGTAATGAATC 67  
Db |||  
QY 1 MetValAlaGlyAlaMetGluAsnArgAspProGlySerGlySerGlyAsnGluVal 20  
Db |||  
QY 68 ATAGAAGGCCCAAAATCAGAGCTCTGAGGGCTCCAGGCTCCGTTCACTGACAC 127  
Db |||  
QY 21 IleGluGlyProGlnAsnAlaArgValLeuLysGlySerGlnAlaArgPheAsnCysThr 40  
QY 128 GTCTCCAGGGTGGAGCTCATCATGCTGGCTCTCAGTGACATGGTGTGCTAAGCGTC 187  
Db |||  
QY 41 ValSerGlnGlyTrpLysLeuIleMetTrpAlaLeuSerAspMetValValLeuSerVal 60  
QY 188 AGGCCCATGGAGCCCATCATCACCAATGACCGCTTCACTCTCAGAGGTACACAGGCG 247  
Db |||  
QY 61 ArgProMetGluProIleIleThrAsnAspArgPheThrSerGlnArgTrpAspGlnGly 80  
QY 248 GGGAACTTCACCTCGGAGATGATCATCCAAATGTGGAGCCGAGTATTCGGGGAATC 307  
Db |||  
QY 81 GlyAsnPheThrSerGluMetIleIleHisAsnValGluProSerAspSerGlyAsnIle 100  
QY 308 AGATCAGGCTCCAGAACAGTCGCTGCATGATCTGCTTACCTTACCGTCCAGATTATG 367  
Db |||  
QY 101 ArgCysSerLeuGlnAsnSerArgLeuHisGlySerAlaTrpLeuThrValGlnValMet 120  
QY 368 GGAGAGCTGTTCAATCCCAAGTCTTAATCTGTAGTGGCTGAGAATCAACCTTGTGAAGTT 427  
Db |||  
QY 121 GlyGluLeuPheIleProSerValAsnLeuValAlaGluAsnGluProCysGluVal 140  
QY 428 ACTTGTCTACCTCAGCTGAGCTGGCTCCGGATATTTCTGGAGTCCGGTCTCTG 487  
Db |||  
QY 141 ThrCysLeuProSerHisTrpThrTrpLeuProAspIleSerTrpGluLeuGlyLeuLeu 160  
QY 488 GTGAGCCATTCAAGCTATTATTTGTTCCGAGCCAGACCTTCAAGTCCAGTGAGC 547  
Db |||  
QY 161 ValSerHisSerSerTrpPheValProGluProSerAspLeuGlnSerAlaValSer 180  
QY 548 ATCTGGTCTGACCCACAGAGCAATGGACTTTGACTTGGCTGCTACCTGGAAGAGC 607  
Db |||  
QY 181 IleLeuAlaLeuThrProGlnSerAsnGlyThrLeuThrCysValAlaThrTrpLysSer 200  
QY 608 CTGAGGCGCCGAGCTGCACTGCAATCTCACTGTGATTCGGTGTCCCAAGACACT 667  
Db |||  
QY 201 LeuLysAlaArgLysSerAlaThrValAsnLeuThrValIleArgCysProGlnAspThr 220  
QY 668 GGAGTGTATTAAATATCCAGGTCTATTATCAAGTTTACCGAGTTAGTGTTCATTG 727  
Db |||  
QY 221 GlyGlyGlyIleAsnIleProGlyValLeuSerSerLeuProSerLeuGlyPheSerLeu 240  
QY 728 CCTACTGGGGAAAGTTGGACTTGGACTAGCAGCCACCATGCTTCTGACGCCGACGTG 787  
Db |||  
QY 241 ProThrTrpGlyLysValGlyLeuGlyLeuAlaGlyThrMetLeuLeuThrProThrCys 260  
QY 788 ACTCTTCAATACGCTGCTGCTGCTGCCGCTGCTGTTGTTGGTGTCACTGCTGCTGC 847  
Db |||  
QY 261 ThrLeuThrIleArgCysCysCysArgArgCysCysGlyCysAsnCysCys 280  
QY 848 CGTTGTTGTTCTGCTGTAGAGAAAGAGAGATTTCGTATTCAATTTCAAAAGAAATCT 907  
Db |||  
QY 281 ArgCysCysPheCysCysArgArgLysArgGlyPheArgIleGlnPheGlnLysLysSer 300  
QY 908 GAAAAGAGAGACAAACAAAGAACTGAGACAGAAAGTGGAAATGAAATCTCCGGCTAC 967  
Db |||  
QY 301 GluLysGluLysThrAsnLysGluThrGluThrGluSerGlyAsnGluAsnSerGlyTrp 320  
QY 968 AATTCAGATGACAAAGACACACACCGCTTCTCCCTCCCAATTCCTGTGAATCC 1027  
Db |||  
QY 321 AsnSerAspGluGlnLysThrThrAspThrAlaSerLeuProProLysSerCysGluSer 340  
QY 1028 AGTGTCTGAAACAAAGAAACAGTAGCTGTGSCCTCTCTCCAGCGGCTGATCAAGT 1087  
Db |||  
QY 341 SerAspProGluGlnArgAsnSerSerCysGlyProProHisGlnArgAlaAspGlnArg 360  
Db |||

QY 1088 CCACCCAGCCAGCAAGTCATCCAGGCTCTTTTAATCTGCAGTCTCTGAGAAGTTC 1147  
Db |||  
QY 361 ProProArgProAlaSerHisProGlnAlaSerPheAsnLeuAlaSerProGluLysVal 380  
Db |||  
QY 1148 AGTAATACAACTGTAGTA 1165  
Db 381 SerAsnThrThrValVal 386  
Db |||  
RESULT 2  
AAU75540  
ID AAU75540 standard; protein; 392 AA.  
XX  
AC AAU75540;  
XX  
DT 23-APR-2002 (first entry)  
XX  
DE Human B7-like protein, B7-L\_h1.  
XX  
KW Human; B7-like protein; B7-L; antiinfertility; gynaecological;  
KW antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;  
KW antiinflammatory; dermatological; antipsoriatic; neuroprotective;  
KW antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;  
KW antiaslathmic; nephrotropic; antibacterial; virucide; tumour; cancer;  
KW reproductive disorder; graft versus host disease; autoimmune disease;  
KW toxic shock syndrome; allergy; nephropathy; skin disorder;  
KW endocrinopathy; lymphoproliferative disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200200710-A2.  
XX  
PD 03-JAN-2002.  
XX  
PF 28-JUN-2001; 2001WO-US020719.  
XX  
PR 28-JUN-2000; 2000US-0214512P.  
PR 28-NOV-2000; 2000US-00729264.  
XX  
PA (AMGE-) AMGEN INC.  
XX  
PI Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;  
XX  
DR WPI; 2002-130881/17.  
DR N-PSDB; ABK13028.  
XX  
PT New B7-like polypeptides, polynucleotides and their modulators, useful  
PT for diagnosing, preventing and treating reproductive, immune and  
PT proliferative disorders, e.g. cancer and arteriosclerosis.  
XX  
PS Claim 13; Fig 1; 135pp; English.  
XX  
CC The invention relates to an isolated B7-like (B7-L) polypeptide (I). The  
CC polypeptide, polynucleotide encoding it and antibody against (I) are  
CC useful for treating B7-like polypeptide-related disease, disorders or  
CC conditions including reproductive disorders (e.g. infertility,  
CC miscarriage, preterm labour and delivery and endometriosis) and  
CC proliferative disorders. Antibodies, soluble proteins comprising  
CC extracellular domains and other regulators of B7-L polypeptides are  
CC useful for enhancing the immune response to tumours. (I) plays a role in  
CC growth and maintenance of cancer cells based on the observation of  
CC seminal vesicle hyperplasia in transgenic mice overexpressing B7-L  
CC polypeptide. Hence modulators of (I) are useful for the treatment of  
CC cancer including seminal vesicle cancer, lung, brain, breast, ovarian,  
CC testicular cancer and cancers of haematopoietic system. B7-L polypeptide  
CC pathway can be manipulated to regulate cytotoxic T-lymphocyte response in  
CC allograft transplantation, graft versus host disease, T-cell dependent B-  
CC cell mediated diseases and autoimmune diseases. B7-L molecules are useful  
CC for alleviating the symptoms associated with diseases involving chronic  
CC immune cell dysfunction or to treat autoimmune diseases such as systemic  
CC lupus erythematosus, rheumatoid arthritis, multiple sclerosis, diabetes,  
CC immune thrombocytopenic purpura and psoriasis, chronic inflammatory  
CC disease such as inflammatory bowel disease (Crohn's disease and  
CC ulcerative colitis), Grave's disease, Hashimoto's thyroiditis and

CC diabetes mellitus. They are also useful as immunosuppressive agents for  
 CC bone marrow and organ transplantation or to prolong graft survival. B7-L  
 CC molecules are also useful for diagnosis and treatment of diseases  
 CC involving abnormal cell proliferation, including arteriosclerosis and  
 CC vascular restenosis. Antagonists of B7-L polypeptides are useful for  
 CC alleviation of toxic shock syndrome or all sensitization due to blood  
 CC transfusions, and for treatment of allergy, asthma and hypersensitivity  
 CC reactions, nephropathies (e.g. glomerulonephritis), skin disorders  
 CC (pemphigus and pemphigoid), endocrinopathies (Grave's disease), various  
 CC pneumopathies (extrinsic alveolitis), vasculopathies, coeliac disease,  
 CC anaemia, thrombocytopenias, Guillain-Barre syndrome and myasthenia  
 CC gravis, and lymphoproliferative disorders such as multiple myeloma. The  
 CC present sequence represents the amino acid sequence of human B7-L<sub>h1</sub>  
 XX  
 SQ Sequence 382 AA;

Alignment Scores:  
 Pred. No.: 1,37e-238 Length: 382  
 Score: 257.00 Matches: 357  
 Percent Similarity: 99.72% Conservative: 0  
 Best Local Similarity: 99.72% Mismatches: 1  
 Query Match: 66.24% Indels: 0  
 DB: 5 Gaps: 0

US-09-729-264-3 (1-1168) x AAU75540 (1-382)

QY 92 GTCTGAAGGGCTCCAGGGCTCGCTTCAACTGCACCGCTCTCCAGGCGTGAAGTCATC 151  
 DB 25 ValLeuLysGlySerGlnAlaArgPheAsnCysThrValSerGlnGlyTrpLysLeuLe 44  
 QY 152 ATGTGGGCTCTCAGTGACATCGTGGTCTAAGCGTCAAGCGCCATGGAGCCCATCATCACC 211  
 DB 45 MetTrpAlaLeuSerAspMetValValLeuSerValArgProMetGluProIleThr 64  
 QY 212 AATGACCGCTTACCTCTCAGAGTACGACGAGCGCGGAACTTCACTCGGAGATCATC 271  
 DB 65 AsnAspArgPheThrSerGlnArgTyrAspGlnGlyAsnPheThrSerGluMetIle 84  
 QY 272 ATCCACAATGTGGAGCCAGTGTTCGGGGAACATCAGATCAGACCTCCAGAACAGTCGC 331  
 DB 85 IleHisValGluProSerAspSerGlyAsnIleArgCysSerLeuGlnAsnSerArg 104  
 QY 332 CTGCATGATCTGCTTACCTTACCGTCCAAAGTTATGGAGAGCTGTTCATTCGCCAGTGT 391  
 DB 105 LeuHisGlySerAlaTyrLeuThrValGlnValMetGlyGluLeuPheIleProSerVal 124  
 QY 392 AATCTTGTAGTCGTGAGATGAACCTTGAGCTTACTTGTCTACCTCAGCTGAGCC 451  
 DB 125 AsnLeuValValAlaGluAsnGluProCysGluValThrCysLeuProSerHisTrpThr 144  
 QY 452 TGGCTCCGGATATTTCCTGGAGCTCGGTCTCGTCTCGCTGAGCCATTCAGAGCTATTATTT 511  
 DB 145 ArgLeuProAspIleSerTrpGluGlyLeuLeuValSerHisSerSerTyrTrpPhe 164  
 QY 512 GTTCCGAGGCCAGCGACCTTCAAAGTCAGTGAAGCATCTGTGCTCTGACCCCAAGAGC 571  
 DB 165 ValProGluProSerAspLeuGlnSerAlaValSerIleLeuAlaLeuThrProGlnSer 184  
 QY 572 AATGGGACTTGTACTCGGTGGTCTACTCGGAGAGCTGAGGCGCCAGCTCTGCAACT 631  
 DB 185 AsnGlyThrLeuThrCysValAlaThrTrpLysSerLeuLysAlaArgLysSerAlaThr 204  
 QY 632 GTAATCTCAGTGTGATTCGGTGTCCCAAGACACTGGAGGTGGTATTAATATTCACAGT 691  
 DB 205 ValAsnLeuThrValIleArgCysProGlnAspThrGlyGlyGlyIleAsnIleProGly 224  
 QY 692 GTATTATCAAGTTACCGAGTTTAGTTTTCATTGCTTACTTGGCGCAAAAGTGGACTT 751  
 DB 225 ValLeuSerSerLeuProSerLeuGlyPheSerLeuProThrTrpGlyLysValGlyLeu 244  
 QY 752 GGACTAGAGCACCATCTTCTGACCGGACGAGTGATCTTACAAATACGCTGTGCTGTC 811  
 DB 245 GlyLeuAlaGlyThrMetLeuLeuThrProThrCysThrLeuThrIleArgCysCys 264

QY 812 TGCCGCGCTGTTGTTGTGTGCTGAACTGCTGCTGCGGTTGTTTCTCTCTGTAGAGA 871  
 DB 265 CysArgArgArgCysCysGlyCysAsnCysCysCysArgCysCysPheCysCysArgArg 284  
 QY 872 AAAAGAGGATTCCTGTTATTCATTTTCAAAGAAATCTGAAAAGAGAGAGACAAAGAA 931  
 DB 285 LysArgGlyPheArgIleGlnPheGlnLysLysSerGluLysGluLysThrAsnLysGlu 304  
 QY 932 ACTGACACAGAAAGTGAATGAAATCCGGCTACAAATTCAGATGAACAAAAGACACCA 991  
 DB 305 ThrGluThrGluSerGlyAsnGluAsnSerGlyTyrAsnSerAspGluGlnLysThrThr 324  
 QY 992 GACACCGCTTCTCTCCCTCCCAATCTGTGAATCCAGTGCATCTGACAAAAGAAACAGT 1051  
 DB 325 AspThrAlaSerLeuProLysSerCysGluSerSerAspProGluGlnArgAsnSer 344  
 QY 1052 AGCTGTGGCGCTCTCTACACAGCGGGCTGATCAAGTCCACCGCCAGGCGCAAGTCATCCA 1111  
 DB 345 SerCysGlyProProHisGlnArgAlaAspGlnArgProProArgProAlaSerHisPro 364  
 QY 1112 CAGGCTCTCTTTTAACTCTGGCCAGTCTCTGAGAAGGTGAGTAAATACAACTGTAGTA 1165  
 DB 365 GlnAlaSerPheAsnLeuAlaSerProGluLysValSerAsnThrThrValVal 382

# RESULT 3

AAU75542

ID AAU75542 standard; protein; 386 AA.

XX AAU75542;

AC AAU75542;

XX 23-APR-2002 (first entry)

XX Human B7-like protein, B7-L<sub>h3</sub>.

XX Human; B7-like protein; B7-L; antiinfertility; gynaecological;

XX antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;

XX antiinflammatory; dermatological; antipsoriatic; neuroprotective;

XX antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;

XX antiaesthetic; nephrotropic; antibacterial; virucide; tumour; cancer;

XX reproductive disorder; graft versus host disease; autoimmune disease;

XX toxic shock syndrome; allergy; nephropathy; skin disorder;

XX endocrinopathy; lymphoproliferative disorder.

XX Homo sapiens.

XX WO200200710-A2.

XX 03-JAN-2002.

XX 28-JUN-2001; 2001WO-US020719.

XX 28-JUN-2000; 2000US-0214512P.

XX 28-NOV-2000; 2000US-00729264.

XX (AMGE-) AMGEN INC.

XX Welcher AA, Samiento UM, Schultz HJ, Chute HT;

XX WPI; 2002-130881/17.

XX N-FSDS; ABK13030.

XX New B7-like polypeptides, polynucleotides and their modulators, useful

XX for diagnosing, preventing and treating reproductive, immune and

XX proliferative disorders, e.g. cancer and arteriosclerosis.

XX Claim 13; Fig 3; 135pp; English.

XX The invention relates to an isolated B7-like (B7-L) polypeptide (I). The

XX polypeptide, polynucleotide encoding it and antibody against (I) are

XX useful for treating B7-like polypeptide-related disease, disorders or

XX conditions including reproductive disorders (e.g. infertility,

XX miscarriage, preterm labour and delivery and endometriosis) and





CC from 1970 fully defined nucleotide sequences which encode novel  
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 CC or its partial peptide, an antibody binding to the polypeptide or peptide  
 CC of the polynucleotide, immunologically assaying the polypeptide or  
 CC peptide of the polynucleotide by contacting the polypeptide or peptide  
 CC with the antibody of the encoded protein, and observing the binding  
 CC between the two, a transformant carrying the polynucleotide in an  
 CC expressible manner and an antisense polynucleotide. The oligonucleotide  
 CC is useful as a primer for synthesizing the polynucleotide, or as a probe  
 CC for detecting the polynucleotide. The polynucleotides and encoded  
 CC proteins are useful as pharmaceutical agents and many disease-related  
 CC genes may be included in them, for developing a diagnostic marker or  
 CC medicines for regulation of their expression and activity, or as targets  
 CC of gene therapy. The genes are involved in tissue and/or cell  
 CC regeneration. Membrane proteins, signal transduction-related proteins,  
 CC transcription-related proteins, disease-related proteins and genes  
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
 CC neurological diseases, cancer, tumours. The cDNA may be used to regulate  
 CC the activity or expression of the encoded protein to treat diseases. The  
 CC sequence presented is a protein of the invention. Note: Some of the  
 CC sequence data for this patent is not represented in the printed  
 CC specification, but is based on sequence information supplied by the  
 CC European Patent Office.

XX SQ Sequence 407 AA;

Alignment Scores:  
 Pred. No.: 1,2e-182 Length: 407  
 Score: 199.00 Matches: 299  
 Percent Similarity: 99.67% Conservative: 0  
 Best Local Similarity: 99.67% Mismatches: 1  
 Query Match: 51.29% Indels: 0  
 DB: 7 Gaps: 0

US-09-729-264-3 (1-1168) x ADB64920 (1-407)

QY 92 GTCCTGAAGGCTCCAGGTCGCTTCAACTGCACCGCTCCAGGCTGGAAGCTCATC 151  
 Db 50 ValLeuLysGlySerGlnAlaArgPheAsnGlyThrValSerGlnGlyTrpLysLeu 69  
 QY 152 ATGTGGCTCTCAGTGACATGGTGGTCTAAGCTCAGGCTCCAGGCTCCAGGCTCATCACC 211  
 Db 70 MetTrpAlaLeuSerAspMetValValLeuSerValArgProMetGluProIleThr 89  
 QY 212 AATGACCGCTTCACTCTCAGAGGTACGACCGAGGCGGAACTTCACTCGAGATGATC 271  
 Db 90 AsnAspArgPheThrSerGlnArgTyrAspGlnGlyAsnPheThrSerGluMetIle 109  
 QY 272 ATCCACAAATGTGGAGCCAGTGTATTCGGGGAACATCAGATGACGCTCCAGAACAGTCGC 331  
 Db 110 IleHisAsnValGluProSerAspSerGlyAsnIleArgCysSerLeuGlnAsnSerArg 129  
 QY 332 CTGCATCGATCGCTTACCTTACCGTCCAGTTATGGAGAGCTGTTTCATTCGCCAGTGT 391  
 Db 130 LeuHisGlySerAlaTyrLeuThrValGlnValMetGlyGluLeuPheIleProSerVal 149  
 QY 392 AATCTTCTAGTCGTGAGATGAACCTTGTGAAGTTACTTGTCTACCTCCACTGGACC 451  
 Db 150 AsnLeuValValAlaGluAsnGluProCysGluValThrCysLeuProSerHisTrpThr 169  
 QY 452 TGGCTCCGGATATTTCTGGAGCTCGGTCTCTCGGTGACCATTCAGCTATATTTT 511  
 Db 170 ArgLeuProAspIleSerTrpGluLeuGlyLeuValSerHisSerSerTyrTyrPhe 189  
 QY 512 GTTCCGGAGCCAGCGACTTCAAGTGCAGTGAGCATCTCGGCTCTGACCCACAGAGC 571  
 Db 190 ValProGluProSerAspLeuGlnSerAlaValSerIleLeuAlaLeuThrProGlnSer 209  
 QY 572 AATGGGACTTTGACTTCGGTGGCTTACCTGGAAGAGCTTGAAGCCCGCAAGCTCGCAACT 631  
 Db 210 AsnGlyThrLeuThrCysValAlaThrTrpLysSerLeuLysAlaArgLysSerAlaThr 229  
 QY 632 GTAAATCTCACTGTGATCGGTGCTCCCAAGACACTGAGGTGGTGTATTAATATCCAGGT 691

Db 230 ValAsnLeuThrValIleArgCysProGlnAspThrGlyGlyIleAsnIleProGly 249  
 QY 692 GTATTATCAAGTTTACCGAGTTTGTAGGTTTTCATTGCTACTTGGGGCAAGTTGCACTT 751  
 Db 250 ValLeuSerSerLeuProSerLeuGlyPheSerLeuProThrTrpGlyLysValGlyLeu 269  
 QY 752 GGACTAGCAGGACCATGCTTCTGACGCGACGCTACTCTTACAATACGCTGCTGCTGC 811  
 Db 270 GlyLeuAlaGlyThrMetLeuLeuThrProThrCysThrLeuThrIleArgCysCys 289  
 QY 812 TGGCGCGCTGTTGTTGGCTCAACTGCTGCTGCGGTTGTTGTTTCTGCTGTAGAGA 871  
 Db 290 CysArgArgArgCysCysGlyCysAsnCysCysArgCysCysPheCysCysArgArg 309  
 QY 872 AAAAGAGGATTTCTGCTTCAATTTCAAAAAGAAATCTGAAAAAGAGAGACAAACAAAGAA 931  
 Db 310 LysArgGlyPheArgIleGlnPheGlnLysLysSerGluLysGluLysThrAsnLysGlu 329  
 QY 932 ACTGAGACAGAAAGTGAATGAAATCCCGCTACAAATTCAGATGAACAAAGACACACA 991  
 Db 330 ThrGluThrGluSerGlyAsnGluAsnSerGlyTyrAsnSerAspGluGlnLysThrThr 349  
 RESULT 5  
 AAU75543  
 ID AAU75543 standard; protein; 377 AA.  
 XX  
 AC AAU75543;  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Human B7-like protein, B7-L\_h4.  
 XX  
 KW Human; B7-like protein; B7-L; antiinfectivity; gynaecological;  
 KW antitumor; cytostatic; immunosuppressive; antiarthritic; antirheumatic;  
 KW antiinflammatory; dermatological; antipsoriatic; neuroprotective;  
 KW antidiabetic; haemostatic; antithyroid; antitumor; antiallergic;  
 KW antiasthmatic; nephrotoxic; antibacterial; virucide; tumour; cancer;  
 KW reproductive disorder; graft versus host disease; autoimmune disease;  
 KW toxic shock syndrome; allergy; nephropathy; skin disorder;  
 KW endocrinopathy; lymphoproliferative disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200710-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 28-JUN-2001; 2001WO-US020719.  
 XX  
 PR 28-JUN-2000; 2000US-0214512P.  
 PR 28-NOV-2000; 2000US-00729264.  
 XX  
 PA (AMGE-) AMGEN INC.  
 XX  
 PI Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;  
 XX  
 DR WPI; 2002-130881/17.  
 DR N-PSDB; ABK13031.  
 XX  
 PT New B7-like polypeptides, polynucleotides and their modulators, useful  
 PT for diagnosing, preventing and treating reproductive, immune and  
 PT proliferative disorders, e.g. cancer and arteriosclerosis.  
 XX  
 PS Claim 13; Fig 4; 135pp; English.  
 XX  
 CC The invention relates to an isolated B7-like (B7-L) polypeptide (I). The  
 CC polypeptide, polynucleotide encoding it and antibody against (I) are  
 CC useful for treating B7-like polypeptide-related disease, disorders or  
 CC conditions including reproductive disorders (e.g. infertility,  
 CC miscarriage, preterm labour and delivery and endometriosis) and  
 CC proliferative disorders. Antibodies, soluble proteins comprising  
 CC extracellular domains and other regulators of B7-L polypeptides are

useful for enhancing the immune response to tumours. (I) plays a role in growth and maintenance of cancer cells based on the observation of seminal vesicle hyperplasia in transgenic mice overexpressing B7-L polypeptide. Hence modulators of (I) are useful for the treatment of cancer including seminal vesicle cancer, lung, brain, breast, ovarian, testicular cancer and cancers of haematopoietic system. B7-L polypeptide pathway can be manipulated to regulate cytotoxic T-lymphocyte response in allograft transplantation, graft versus host disease, T-cell dependent B cell mediated diseases and autoimmune diseases. B7-L molecules are useful for alleviating the symptoms associated with diseases involving chronic immune cell dysfunction or to treat autoimmune diseases such as systemic lupus erythematosus, rheumatoid arthritis, multiple sclerosis, diabetes, immune thrombocytopenic purpura and psoriasis, chronic inflammatory disease such as inflammatory bowel disease (Crohn's disease and ulcerative colitis). Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They are also useful as immunosuppressive agents for bone marrow and organ transplantation or to prolong graft survival. B7-L molecules are also useful for diagnosis and treatment of diseases involving abnormal cell proliferation, including arteriosclerosis and vascular restenosis. Antagonists of B7-L polypeptides are useful for alleviation of toxic shock syndrome or allosensitisation due to blood transfusions, and for treatment of allergy, asthma and hypersensitivity reactions, nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and pemphigoid), endocrinopathies (Grave's disease), various pneumopathies (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia, thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and lymphoproliferative disorders such as multiple myeloma. The present sequence represents the amino acid sequence of human B7-L<sub>h4</sub>

Sequence 377 AA;

RESULT 6  
ABG28169  
ID ABG28169 standard; protein; 463 AA.  
XX AC ABG28169;  
XX AC ABG28169;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #28160.  
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX KW  
XX OS Homo sapiens.  
XX OS WO200175067-A2.  
XX PN 11-OCT-2001.  
XX PD  
XX PF 30-MAR-2001; 2001WO-US008631.  
XX PF 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.  
XX XX  
XX PA (HYSE-) HYSEQ INC.  
XX PI Dmanac RT, Liu C, Tang YT;  
XX PI WPI: 2001-639362/73.  
XX DR N-PSDB; AAS922356.  
XX DR  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX PT  
XX PS Claim 20; SEQ ID NO 58528; 103pp; English.  
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX CC  
XX SQ Sequence 463 AA;

Alignment Scores:			
Pred. No.:	1_12e-127	Length:	463
Score:	142.00	Matches:	142
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	36.60%	Indels:	0
DB:	4	Gaps:	0

US-09-729-264-3 (1-1169) x ABG28169 (1-463)

QY 455 CTCGCCGGATTTTCCTCGGAGCTCGGTCTCCTGGTACGCATTCAAGCTAATTTTGTT 514

Db 102 LeuProAspIleSerTrpGluLeuGlyLeuLeuValSerHisSerTyrTyrPheVal 121  
QY 515 CCGAGCCGACGACCTTCAAGTGCAGTGCAGCATCTGCTGACCCACACAGCAAT 574  
Db 122 ProGluProSerAspLeuGlnSerAlaValSerIleLeuAlaLeuThrProGlnSerAsn 141  
QY 575 GGGACTTTGACTTGGCTGCTACTGGAAGAGCTGAAGGCCGCAAGTCTGCAACTGTA 634  
Db 142 GlyThrLeuThrCysValAlaThrTrpLysSerLeuLysAlaArgLysSerAlaThrVal 161  
QY 635 AATCTCACTGTGATTCGGTGTCCCAAGACACATGGAGTGGTATTAATATTCAGGGTGA 694  
Db 162 AsnLeuThrValIleArgCysProGlnAspThrGlyGlyIleAsnIleProGlyVal 181  
QY 695 TTATCAAGTTTACCGAGTTTGGTGTCTTCTGCTACTTGGGCAAGTTGGACTTGGGA 754  
Db 182 LeuSerSerLeuProSerLeuGlyPheSerLeuProThrTrpGlyLysValGlyLeuGly 201  
QY 755 CTAGCAGGCACCATGCTTCTGAGCCGACGCTGACTCTTACAATACGCTGCTGCTGCTGC 814  
Db 202 LeuAlaGlyThrMetLeuLeuThrProThrCysThrLeuThrIleArgCysCysCys 221  
QY 815 CGCGCTGTTGTTGGTGCCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 874  
Db 222 ArgArgArgCysCysGlyCysAsnCysCysArgCysCysPheCysCysArgArgLys 241  
QY 875 AGAGGA 880  
Db 242 ArgGly 243

RESULT 7  
AAM20679  
ID AAM20679 standard; protein; 78 AA.  
AC AAM20679;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Peptide #7113 encoded by probe for measuring cervical gene expression.  
XX  
XX Probe; human; microarray; gene expression; cervical epithelial cell;  
XX cervical cancer.  
XX Homo sapiens.  
XX  
XX WO200157278-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000670.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488901/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human cervical epithelial cells.  
XX  
XX Claim 27; SEQ ID NO 25505; 487pp; English.  
XX  
XX The present invention relates to human single exon nucleic acid probes  
XX (SENPs; see AAI10068-AA128459). The present sequence is a peptide encoded

CC by one such probe. The SNPs are derived from human HeLa cells. The SNPs  
CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 78 AA;  
Alignment Scores:  
Pred. No.: 0.0327 Length: 78  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.17% Indels: 0  
DB: 4 Gaps: 0  
US-09-729-264-3 (1-1168) x AAM20679 (1-78)  
QY 835 GCAGCCACAAACGACGCGCAGCAGCAGCAGCAGCG 800  
Db 4 AlaAlaThrThrThrAlaAlaAlaAlaAla 15  
RESULT 8  
ABB42042  
ID ABB42042 standard; peptide; 78 AA.  
XX  
XX AC ABB42042;  
XX  
XX DT 04-FEB-2002 (first entry)  
XX  
XX DE Peptide #9548 encoded by human foetal liver single exon probe.  
XX  
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200157277-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 30-JAN-2001; 2001WO-US000669.  
XX  
XX PR 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human foetal liver.  
XX  
XX Claim 27; SEQ ID NO 34677; 639pp + Sequence Listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for measuring  
XX human gene expression in a sample derived from human foetal liver. The  
XX single exon nucleic acid probes may be used for predicting, measuring and  
XX displaying gene expression in samples derived from human foetal liver. The  
XX present sequence is a peptide encoded by a single exon nucleic acid probe  
XX of the invention. Note: The sequence data for this patent did not form  
XX part of the printed specification, but was obtained in electronic format  
XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 78 AA;

Alignment Scores:  
 Pred. No.: 0.0327 Length: 78  
 Score: 12.00 Matches: 12  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.17% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x ABB42042 (1-78)

QY 835 GCAGCCACAAACACGCGCGGCGAGCAGCAGCG 800  
 DB 4 AlaAlaThrThrThrAlaAlaAlaAlaAla 15

RESULT 9  
 AAM35844  
 ID AAM35844 standard; protein; 78 AA.  
 XX AAM35844;  
 AC AAM35844;  
 XX  
 DT 17-OCT-2001 (first entry)  
 XX  
 DE Peptide #9881 encoded by probe for measuring placental gene expression.  
 XX  
 KW Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157272-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000663.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488897/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human placenta.  
 XX  
 PS Claim 27; SEQ ID NO 36113; 654pp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes (SENPs;  
 CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for producing a microarray for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from human placenta. The probes are useful for antenatal diagnosis of  
 CC human genetic disorders

SQ Sequence 78 AA;

Alignment Scores:  
 Pred. No.: 0.0327 Length: 78  
 Score: 12.00 Matches: 12  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.17% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x ABB25654 (1-78)

QY 835 GCAGCCACAAACACGCGCGGCGAGCAGCAGCG 800  
 DB 4 AlaAlaThrThrThrAlaAlaAlaAlaAla 15

RESULT 10  
 ABB25654  
 ID ABB25654 standard; protein; 78 AA.  
 XX ABB25654;  
 AC ABB25654;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Protein #7653 encoded by probe for measuring heart cell gene expression.  
 XX  
 KW Human; gene expression; heart; microarray; vascular system;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157274-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000666.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488899/53.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts.  
 XX  
 PS Claim 15; SEQ ID NO 27424; 530pp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart (see  
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
 CC probe. The probes may be used for predicting, measuring and displaying  
 CC gene expression in samples derived from the human heart via microarrays.  
 CC By measuring gene expression, the probes are useful for predicting,  
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
 CC human heart and vascular system e.g. cardiovascular disease,  
 CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 78 AA;

Alignment Scores:  
 Pred. No.: 0.0327 Length: 78  
 Score: 12.00 Matches: 12  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.17% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x ABB25654 (1-78)

QY 835 GCAGCCACAAACACGCGCGGCGAGCAGCAGCG 800

```
Db      4 AlaAlaThrThrThrAlaAlaAlaAlaAlaAla 15
RESULT 11
AAM75735
ID      AAM75735 standard; protein; 78 AA.
XX      AC      AAM75735;
XX      DT      06-NOV-2001 (first entry)
XX      DE      Human bone marrow expressed probe encoded protein SEQ ID NO: 36041.
XX      KW      Human; bone marrow expressed exon; gene expression analysis; probe;
XX      KW      microarray; cancer; leukaemia; lymphoma; myeloma.
XX      OS      Homo sapiens.
XX      PN      WO200157276-A2.
XX      PD      09-AUG-2001.
XX      PF      30-JAN-2001; 2001WO-US000668.
XX      PR      04-FEB-2000; 2000US-0180312P.
XX      PR      26-MAY-2000; 2000US-0207456P.
XX      PR      30-JUN-2000; 2000US-00608408.
XX      PR      03-AUG-2000; 2000US-00632366.
XX      PR      21-SEP-2000; 2000US-0234687P.
XX      PR      27-SEP-2000; 2000US-0236359P.
XX      PR      04-OCT-2000; 2000GB-00024263.
XX      PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX      PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX      DR      WPI; 2001-488900/53.
XX      PT      Human genome-derived single exon nucleic acid probes useful for analyzing
XX      PT      gene expression in human bone marrow.
XX      PS      Example 4; SEQ ID NO 36041; 658pp + Sequence Listing; English.
XX      CC      The present invention provides a number of single exon nucleic acid
XX      CC      probes which are derived from genomic sequences expressed in the human
XX      CC      bone marrow. They can be used to measure gene expression in bone marrow
XX      CC      samples, which may enable the improved diagnosis and treatment of cancers
XX      CC      such as lymphoma, leukaemia and myeloma. The present sequence is a
XX      CC      protein encoded by one of the probes of the invention
XX      SQ      Sequence 78 AA;

Alignment Scores:
Pred. No.:      0.0327      Length:      78
Score:          12.00      Matches:      12
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    3.17%      Indels:      0
DB:             4      Gaps:      0

US-09-729-264-3 (1-1168) x AAM75735 (1-78)
QY      835 GCAGCCACACAGCGCGGCGAGCAGCAGCG 800
Db      4 AlaAlaThrThrThrAlaAlaAlaAlaAlaAla 15
RESULT 12
AAM62923
ID      AAM62923 standard; protein; 78 AA.
XX      AC      AAM62923;
XX      DT      05-NOV-2001 (first entry)
```

```
XX      Human brain expressed single exon probe encoded protein SEQ ID NO: 35028.
XX      DE      Human; brain expressed exon; gene expression analysis; probe; microarray;
XX      KW      Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX      OS      Homo sapiens.
XX      PN      WO200157275-A2.
XX      PD      09-AUG-2001.
XX      PF      30-JAN-2001; 2001WO-US000667.
XX      PR      04-FEB-2000; 2000US-0180312P.
XX      PR      26-MAY-2000; 2000US-0207456P.
XX      PR      30-JUN-2000; 2000US-00608408.
XX      PR      03-AUG-2000; 2000US-00632366.
XX      PR      21-SEP-2000; 2000US-0234687P.
XX      PR      27-SEP-2000; 2000US-0236359P.
XX      PR      04-OCT-2000; 2000GB-00024263.
XX      PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX      PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX      DR      WPI; 2001-483446/52.
XX      PT      Single exon nucleic acid probes for analyzing gene expression in human
XX      PT      brains.
XX      PS      Example 4; SEQ ID NO 35028; 650pp + Sequence Listing; English.
XX      CC      The present invention provides a number of single exon nucleic acid
XX      CC      probes which are derived from genomic sequences expressed in the human
XX      CC      brain. They can be used to measure gene expression in brain cell samples,
XX      CC      which may enable the diagnosis and improved treatment of nervous system
XX      CC      diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX      CC      epilepsy and cancers. The present sequence is a protein encoded by one of
XX      CC      the probes of the invention
XX      SQ      Sequence 78 AA;

Alignment Scores:
Pred. No.:      0.0327      Length:      78
Score:          12.00      Matches:      12
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    3.17%      Indels:      0
DB:             4      Gaps:      0

US-09-729-264-3 (1-1168) x AAM62923 (1-78)
QY      835 GCAGCCACACAGCGCGGCGAGCAGCAGCG 800
Db      4 AlaAlaThrThrThrAlaAlaAlaAlaAlaAla 15
RESULT 13
ABG57473
ID      ABG57473 standard; peptide; 78 AA.
XX      AC      ABG57473;
XX      DT      25-FEB-2003 (first entry)
XX      DE      Human liver peptide, SEQ ID No 36121.
XX      KW      Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX      KW      hypercholesterolaemia; coronary heart disease.
XX      OS      Homo sapiens.
XX      PN      WO200157273-A2.
```

XX 09-AUG-2001.  
 XX  
 XX PF 30-JAN-2001; 2001WO-US000664.  
 XX  
 XX PR 04-FEB-2000; 2000US-0180312P.  
 XX PR 26-MAY-2000; 2000US-0207456P.  
 XX PR 30-JUN-2000; 2000US-00608408.  
 XX PR 03-AUG-2000; 2000US-00632366.  
 XX PR 21-SEP-2000; 2000US-0234687P.  
 XX PR 27-SEP-2000; 2000US-0236359P.  
 XX PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 XX WPI; 2001-488898/53.  
 XX  
 XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human adult liver.  
 PT  
 PS Claim 27; SEQ ID NO 36121; 658pp; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridizes at high  
 CC stringency to a nucleic acid molecule expressed in the human adult liver.  
 CC (I) may be used for predicting, measuring and displaying gene expression  
 CC in samples derived from human adult liver. The genes identified may be  
 CC involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinemia, hyperlipidaemia and hypercholesterolaemia which is  
 CC associated with coronary heart disease. ABG47348-ABG5930 represent human  
 CC liver single exon encoded peptides of the invention. Note: The sequence  
 CC information for this patent does not appear in the printed specification  
 CC but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 78 AA;  
 Alignment Scores:  
 Pred. No.: 0.0327 Length: 78  
 Score: 12.00 Matches: 12  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.17% Indels: 0  
 DB: 4 Gaps: 0  
 US-09-729-264-3 (1-1168) x ABG57473 (1-78)  
 Qy 835 GCAGCCACACACGACGCGGACGACGACGAGCG 800  
 Db 4 AAlaAlaThrThrThrAlaAlaAlaAlaAla 15  
 RESULT 14  
 ABG45219  
 ID ABG45219 standard; peptide; 78 AA.  
 XX  
 AC ABG45219;  
 XX  
 DT 19-AUG-2002 (first entry)  
 XX  
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 34884.  
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberculous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;

KW hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 XX 30-JAN-2001; 2001WO-US000665.  
 XX  
 XX 04-FEB-2000; 2000US-0180312P.  
 XX PR 26-MAY-2000; 2000US-0207456P.  
 XX PR 30-JUN-2000; 2000US-00608408.  
 XX PR 03-AUG-2000; 2000US-00632366.  
 XX PR 21-SEP-2000; 2000US-0234687P.  
 XX PR 27-SEP-2000; 2000US-0236359P.  
 XX PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 XX WPI; 2002-114183/15.  
 DR  
 XX  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples.  
 PT  
 XX Claim 27; SEQ ID NO 34884; 634pp; English.  
 PS  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of probes  
 CC; the novel set of probes which hybridise at high stringency to a nucleic  
 CC acid expressed in the human lung; measuring gene expression in a sample  
 CC derived from human lung, comprising (a) contacting the array with a  
 CC collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of the  
 CC array; identifying exons in a eukaryotic genome, comprising (a)  
 CC algorithmically predicting at least one exon from genomic sequences of  
 CC the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene expression  
 CC analysis, and for identifying exons in a gene, particularly using human  
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 CC tuberculous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
 CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 CC present sequence is a peptide/protein encoded by a single exon probe of  
 CC the invention. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 78 AA;  
 Alignment Scores:  
 Pred. No.: 0.0327 Length: 78  
 Score: 12.00 Matches: 12



Percent Similarity: 100.00%    Conservative: 0  
Best Local Similarity: 100.00%    Mismatches: 0  
Query Match: 3.17%    Indels: 0  
DB: 5    Gaps: 0

US-09-729-264-3 (1-1168) x ABG45219 (1-78)

QY 835 GCGCCACACACGACGCGCGGCGGCGAGCAGCAGCG 800  
Db 4 AlaAlaThrThrThrAlaAlaAlaAlaAlaAla 15

RESULT 15

AAU07370  
ID AAU07370 standard; protein; 524 AA.

XX AC AAU07370;

XX DT 18-DEC-2001 (first entry)

XX DE G protein-coupled receptor.

XX KW Human; mental disorder; thyroid disease; renal failure; anorexia;  
KW inflammatory condition; Crohn's disease; rheumatoid arthritis; HIV;  
KW autoimmune disorder; schizophrenia; migraine; stroke; dementia; obesity;  
KW depression; Parkinson's disease; Alzheimer's disease; viral infection;  
KW Huntington's disease; human immunodeficiency virus; type 2 diabetes;  
KW anorexia; hypotension; hypertension; thrombosis; myocardial infarction;  
KW atherosclerosis; cancer; sexual dysfunction; G protein-coupled receptor;  
KW nGPCR.  
XX OS Homo sapiens.  
XX PN WO200162924-A2.  
XX PD 30-AUG-2001.  
XX PF 23-FEB-2001; 2001WO-US005989.  
XX PR 24-FEB-2000; 2000US-0184602P.  
XX PR 24-FEB-2000; 2000US-0184604P.  
XX PR 24-FEB-2000; 2000US-0184606P.  
XX PR 24-FEB-2000; 2000US-0184689P.  
XX PR 24-FEB-2000; 2000US-0184690P.  
XX PR 24-FEB-2000; 2000US-0184710P.  
XX PR 24-FEB-2000; 2000US-0184712P.  
XX PR 24-FEB-2000; 2000US-0184715P.  
XX PR 24-FEB-2000; 2000US-0184716P.  
XX PR 24-FEB-2000; 2000US-0184725P.  
XX PR 24-FEB-2000; 2000US-0184822P.  
XX FA (PHAA ) PHARMACIA & UPJOHN CO.  
XX PI Vogeli G, Wood LS, Parodi LA, Lind P;  
XX WPI; 2001-570632/64.  
XX DR Novel nucleic acid and encoded nGPCR-x, used to screen for compounds for  
PT use in the treatment of mental disorders, such as Alzheimer's disease, or  
PT Parkinson's disease.  
XX PS Claim 31; Page 178-179; 263pp; English.

XX CC The invention relates to novel isolated human G protein-coupled receptors  
CC (nGPCR-x). The nGPCR-x can be used for screening compounds which can be  
CC used to treat mental disorders, thyroid disease, renal failure,  
CC inflammatory conditions such as Crohn's disease, rheumatoid arthritis,  
CC autoimmune disorders, schizophrenia, migraine, stroke, dementia,  
CC depression, Parkinson's disease, Alzheimer's disease, and Huntington's  
CC disease. They may also be used for treating viral infections such as  
CC human immunodeficiency virus (HIV), type 2 diabetes, obesity, anorexia,  
CC hypotension, hypertension, thrombosis, myocardial infarction,  
CC atherosclerosis, cancer, and sexual dysfunction. AAU25617-AAU25726  
CC represent the amino acid sequences of novel human G protein-coupled

CC receptors, nGPCR-2031 to nGPCR-2140 respectively, as described in the  
CC invention  
XX SQ Sequence 524 AA;

Alignment Scores:  
Pred. No.: 0.251    Length: 524  
Score: 11.00    Matches: 11  
Percent Similarity: 100.00%    Conservative: 0  
Best Local Similarity: 100.00%    Mismatches: 0  
Query Match: 2.91%    Indels: 0  
DB: 4    Gaps: 0

US-09-729-264-3 (1-1168) x AAU07370 (1-524)

QY 832 GCCACACACGACGCGCGGCGGCGAGCAGCAGCG 800

Db 432 AlaThrThrThrAlaAlaAlaAlaAlaAla 442

RESULT 16

ABG65262

ID ABG65262 standard; protein; 1805 AA.

XX AC ABG65262;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 22578.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US009231.  
XX PR 23-MAR-2000; 2000US-0191637P.  
XX PR 11-JUL-2000; 2000US-00614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX DR N-PSDB; ABL09365.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX PS Disclosure; SEQ ID NO 22578; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 1805 AA;

Alignment Scores:

Pred. No.: 0.223    Length: 1805  
Score: 11.00    Matches: 11  
Percent Similarity: 100.00%    Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.84% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x ABB65262 (1-1805)

QY 801 GGTGTCGTGTCGCCCGCTCTGTTGTGGCT 833  
 DB 27 AlaAlaAlaAlaAlaValValValAla 37

RESULT 17

AAAY65872

ID AAY65872 standard; peptide; 85 AA.

XX AC AAY65872;

XX DT 10-FEB-2000 (first entry)

XX DE n-myc mutant peptide 1.

XX KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;

XX KW mitein.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO9958552-A2.

XX PD 18-NOV-1999.

XX PF 03-MAY-1999; 99WO-NO000143.

XX PR 08-MAY-1998; 98NO-00002097.

XX PA (NHVD ) NORSE HYDRO AS.

PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;

XX WPI; 2000-039064/03.

DR New peptides derived from genes with frameshift mutations, used to

PT develop products for the treatment and prophylaxis of cancers.

XX Claim 13; Page 29; 166pp; English.

XX Peptides AAY65872-Y66142 are fragments of mutant proteins arising from a

CC frameshift mutation in a gene from a cancer cell. The peptides are

CC characterised in that they: (i) are at least 8 amino acids long and a

CC fragment of a mutant protein arising from a frameshift mutation in a gene

CC of a cancer cell; (ii) consist of at least one amino acid of the mutant

CC part of a protein sequence encoded by the gene; (iii) comprise 0-10 amino

CC acid from the carboxyl terminus of the normal part of the protein

CC sequence preceding the amino terminus of the mutant sequence and may

CC further extend to the carboxyl terminus of the mutant part of the protein

CC as determined by a new stop codon generated by the frameshift mutation;

CC and (iv) induce, either in their full lengths or after processing by an

CC antigen presenting cell (APC), T cell responses. The genes that the

CC peptides are derived from, are characterised as susceptible to frameshift

CC mutation by having a mono nucleoside base repeat sequence of at least 5

CC residues, or a di-nucleoside base repeat sequence of at least 4 di-

CC nucleoside base units. The peptides are created by the addition or

CC deletion of 1 or 2 nucleoside base residues from the repeat sequence. The

CC novel peptides can elicit T cell responses and toxicity against tumours

CC and cancer cells carrying genes with frameshift mutations. The novel

CC peptides and DNA sequences can be used for the preparation of a

CC composition for the treatment or prophylaxis of cancer

XX SQ Sequence 85 AA;

Alignment Scores:

Pred. No.: 2.75 Length: 85

Score: 10.00 Matches: 10

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.58% Indels: 0  
 DB: 3 Gaps: 0

US-09-729-264-3 (1-1168) x AAY65872 (1-85)

QY 117 TCAACTGCACCGCTCTCCAGGGCTGAAGC 146

DB 25 SerThrAlaProSerProArgAlaGlySer 34

RESULT 18

AAAY65874

ID AAY65874 standard; peptide; 94 AA.

XX AC AAY65874;

XX DT 10-FEB-2000 (first entry)

XX DE n-myc mutant peptide 3.

XX KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;

XX KW mitein.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO9958552-A2.

XX PD 18-NOV-1999.

XX PF 03-MAY-1999; 99WO-NO000143.

XX PR 08-MAY-1998; 98NO-00002097.

XX PA (NHVD ) NORSE HYDRO AS.

PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;

XX WPI; 2000-039064/03.

DR New peptides derived from genes with frameshift mutations, used to

PT develop products for the treatment and prophylaxis of cancers.

XX Claim 13; Page 29; 166pp; English.

XX Peptides AAY65874-Y66142 are fragments of mutant proteins arising from a

CC frameshift mutation in a gene from a cancer cell. The peptides are

CC characterised in that they: (i) are at least 8 amino acids long and a

CC fragment of a mutant protein arising from a frameshift mutation in a gene

CC of a cancer cell; (ii) consist of at least one amino acid of the mutant

CC part of a protein sequence encoded by the gene; (iii) comprise 0-10 amino

CC acid from the carboxyl terminus of the normal part of the protein

CC sequence preceding the amino terminus of the mutant sequence and may

CC further extend to the carboxyl terminus of the mutant part of the protein

CC as determined by a new stop codon generated by the frameshift mutation;

CC and (iv) induce, either in their full lengths or after processing by an

CC antigen presenting cell (APC), T cell responses. The genes that the

CC peptides are derived from, are characterised as susceptible to frameshift

CC mutation by having a mono nucleoside base repeat sequence of at least 5

CC residues, or a di-nucleoside base repeat sequence of at least 4 di-

CC nucleoside base units. The peptides are created by the addition or

CC deletion of 1 or 2 nucleoside base residues from the repeat sequence. The

CC novel peptides can elicit T cell responses and toxicity against tumours

CC and cancer cells carrying genes with frameshift mutations. The novel

CC peptides and DNA sequences can be used for the preparation of a

CC composition for the treatment or prophylaxis of cancer

XX SQ Sequence 94 AA;

Alignment Scores:

Pred. No.: 2.72 Length: 94

Score: 10.00 Matches: 10

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.58% Indels: 0  
 DB: 3 Gaps: 0

US-09-729-264-3 (1-1168) x AAY65874 (1-94)

QY 117 TCAACTGCACCGTCTCCAGGGCTGGAAGC 146  
 Db 34 SerThrAlaProSerProArgAlaGlySer 43

RESULT 19

AAAY65875  
 ID AAY65875 standard; peptide; 95 AA.

XX AC AAY65875;

XX DT 10-FEB-2000 (first entry)

XX DE n-myc mutant peptide 4.

XX KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;  
 KW mutain.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO9558552-A2.

XX PD 18-NOV-1999.

XX PF 03-MAY-1999; 99WO-NO000143.

XX PR 08-MAY-1998; 98NO-00002097.

XX PA (NH2D) NORSK HYDRO AS.

XX PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;

XX WPI; 2000-039064/03.

XX PT New peptides derived from genes with frameshift mutations, used to  
 PT develop products for the treatment and prophylaxis of cancers.

XX PS Claim 13; Page 29; 166pp; English.

XX CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a  
 CC frameshift mutation in a gene from a cancer cell. The peptides are  
 CC characterised in that they: (i) are at least 8 amino acids long and a  
 CC fragment of a mutant protein arising from a frameshift mutation in a gene  
 CC of a cancer cell; (ii) consist of at least one amino acid of the mutant  
 CC part of a protein sequence encoded by the gene; (iii) comprise 0-10 amino  
 CC acid from the carboxyl terminus of the normal part of the protein  
 CC sequence preceding the amino terminus of the mutant sequence and may  
 CC further extend to the carboxyl terminus of the mutant part of the protein  
 CC as determined by a new stop codon generated by the frameshift mutation;  
 CC and (iv) induce, either in their full lengths or after processing by an  
 CC antigen presenting cell (APC), T cell responses. The genes that the  
 CC peptides are derived from, are characterised as susceptible to frameshift  
 CC mutation by having a mono nucleoside base repeat sequence of at least 5  
 CC residues, or a di-nucleoside base repeat sequence of at least 4 di-  
 CC nucleoside base units. The peptides are created by the addition or  
 CC deletion of 1 or 2 nucleoside base residues from the repeat sequence. The  
 CC novel peptides can elicit T cell responses and toxicity against tumours  
 CC and cancer cells carrying genes with frameshift mutations. The novel  
 CC peptides and DNA sequences can be used for the preparation of a  
 CC composition for the treatment or prophylaxis of cancer  
 XX SQ Sequence 95 AA;

Alignment Scores:

Pred. No.: 2.72 Length: 95  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.58% Indels: 0  
 DB: 3 Gaps: 0

US-09-729-264-3 (1-1168) x AAY65875 (1-95)

QY 117 TCAACTGCACCGTCTCCAGGGCTGGAAGC 146  
 Db 35 SerThrAlaProSerProArgAlaGlySer 44

RESULT 20

ABB64563  
 ID ABB64563 standard; protein; 2168 AA.

XX AC ABB64563;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 20481.

XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX DR N-PSDB; ABL08666.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

XX PS Disclosure; SEQ ID NO 20481; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABBS72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 2168 AA;

Alignment Scores:

Pred. No.: 2.02 Length: 2168  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.65% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x ABB64563 (1-2168)

QY 829 ACAACACGACGCGGCGAGCAGCAGCAGCG 800

Db 973 ThrThrThrAlaAlaAlaAlaAlaAla 982

RESULT 21  
ABB69039  
ID ABB69039 standard; protein; 96 AA.  
XX  
AC ABB69039;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 33909.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI: 2001-656860/75.  
DR N-PSDB; ABLI3142.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 33909; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABLI6176-ABL30511), expressed DNA  
CC sequences (ABLI01840-ABLI6175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 96 AA;  
XX  
Alignment Scores:  
Pred. No.: 25.1 Length: 96  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.38% Indels: 0  
DB: 4 Gaps: 0  
US-09-729-264-3 (1-1168) x ABB69039 (1-96)  
QY 826 ACAACGCGGCGCAGCAGCAGCG 800  
DB 20 ThrThrThrAlaAlaAlaAlaAla 28  
RESULT 22  
ABB63572  
ID ABB63572 standard; protein; 191 AA.  
XX  
AC ABB63572;  
XX  
DT 26-MAR-2002 (first entry)  
XX

DE Drosophila melanogaster polypeptide SEQ ID NO 17508.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI: 2001-656860/75.  
DR N-PSDB; ABL07675.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 17508; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABLI6176-ABL30511), expressed DNA  
CC sequences (ABLI01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 191 AA;  
XX  
Alignment Scores:  
Pred. No.: 23.5 Length: 191  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.38% Indels: 0  
DB: 4 Gaps: 0  
US-09-729-264-3 (1-1168) x ABB63572 (1-191)  
QY 856 ACAACACGCGCAGCAGCAGTTCGCGCC 830  
DB 96 ThrThrThrAlaAlaAlaAlaAla 104  
RESULT 23  
ABB70426  
ID ABB70426 standard; protein; 191 AA.  
XX  
AC ABB70426;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 38070.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX

PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL14529.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
XX Disclosure; SEQ ID NO 38070; 21pp + Sequence Listing; English.  
PS  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 191 AA;  
SQ  
Alignment Scores:  
Pred. No.: 23.5 Length: 191  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.38% Indels: 0  
DB: 4 Gaps: 0  
US-09-729-264-3 (1-1168) x ABB70426 (1-191)  
QY 835 GCAGCCACAAACGACGGCGCAGCA 809  
Db 89 AlaAlaThrThrThrAlaAlaAla 97  
RESULT 24  
ABB67144  
ID ABB67144 standard; protein; 191 AA.  
XX  
XX AC ABB67144;  
XX  
XX DT 26-MAR-2002 (first entry)  
XX  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 28224.  
XX  
XX DE Drosophila; developmental biology; cell signalling; insecticide;  
XX  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX  
XX KW pharmaceutical.  
XX  
XX OS Drosophila melanogaster.  
XX  
XX PN WO200171042-A2.  
XX  
XX PD 27-SEP-2001.  
XX  
XX PF 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX PA (PEKE ) PE CORP NY.  
XX

PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL11247.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
XX Disclosure; SEQ ID NO 28224; 21pp + Sequence Listing; English.  
PS  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 191 AA;  
SQ  
Alignment Scores:  
Pred. No.: 23.5 Length: 191  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.38% Indels: 0  
DB: 4 Gaps: 0  
US-09-729-264-3 (1-1168) x ABB67144 (1-191)  
QY 856 ACAACACGACGACGACGAGTTCAGCC 830  
Db 96 ThrThrThrAlaAlaAlaAlaAla 104  
RESULT 25  
ADA48420  
ID ADA48420 standard; protein; 253 AA.  
XX  
XX AC ADA48420;  
XX  
XX DT 20-NOV-2003 (first entry)  
XX  
XX DE Rice protein conferring disease resistance in plants.  
XX  
XX DE disease resistance; pathogen tolerance; plant pathogen; plant; rice.  
XX  
XX OS Oryza sativa.  
XX  
XX PN WO2003000906-A2.  
XX  
XX PD 03-JAN-2003.  
XX  
XX PF 21-JUN-2002; 2002WO-IB002453.  
XX  
XX PF 22-JUN-2001; 2001US-0300112P.  
PR 26-SEP-2001; 2001US-0352277P.  
PR 22-MAR-2002; 2002US-0366535P.  
XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Glazebrook J, Briggs S, Cooper B, Goff SA, Moughamer T;  
PI Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;  
XX  
XX WPI; 2003-184052/18.  
DR N-ESDB; ADA48419.  
XX  
XX New polynucleotide comprising a plant nucleotide sequence having an open  
PT reading frame that encodes a polypeptide associated with disease  
PT resistance, useful for conferring resistance or tolerance to a plant

PT pathogen.  
PS Claim 10; SEQ ID NO 490; 299pp; English.  
XX  
CC The invention relates to a novel isolated polynucleotide comprising a  
CC plant nucleotide sequence having an open reading frame that encodes a  
CC polypeptide associated with disease resistance or its fragment having  
CC substantially the same activity as the full-length polypeptide. The  
CC polynucleotide of the invention is useful for conferring resistance or  
CC tolerance to a plant pathogen. The present sequence represents a protein  
CC conferring disease resistance used in the invention.  
XX  
SQ Sequence 253 AA;  
  
Alignment Scores:  
Pred. No.: 22.8 Length: 253  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.32% Indels: 0  
DB: 6 Gaps: 0  
  
US-09-729-264-3 (1-1168) x ADA48420 (1-253)  
QY 813 GCCGCGCTGTGTGTGCTGCAACT 839  
DB 36 AlAlaValValValValAlaAlaThr 44  
|||||  
RESULT 26  
ABU17237  
ID ABU17237 standard; protein; 258 AA.  
XX  
AC ABU17237;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #2764.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Acinetobacter baumannii.  
XX  
FN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
Wang L, Zamudio C, Malone C, Haselbeck R, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA21107.  
XX  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 45161; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC the gene product or that has an activity against a biological pathway;  
CC the gene product or that has an activity against a biological pathway;  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 258 AA;  
  
Alignment Scores:  
Pred. No.: 22.8 Length: 258  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.32% Indels: 0  
DB: 6 Gaps: 0  
  
US-09-729-264-3 (1-1168) x ABU17237 (1-258)  
QY 1050 GTAGTGTGGCGCTCTCACAGCGG 1076  
DB 6 ValAlaValAlaLeuLeuThrSergly 14  
|||||  
RESULT 27  
ADA36602  
ID ADA36602 standard; protein; 269 AA.  
XX  
AC ADA36602;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Acinetobacter baumannii protein #3763.  
XX  
KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;  
KW plant biocontrol agent.  
XX  
OS Acinetobacter baumannii.  
XX  
FN US6562958-B1.  
XX  
PD 13-MAY-2003.  
XX  
PF 04-JUN-1999; 99US-00328352.  
XX  
PR 09-JUN-1998; 98US-0088701P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Breton G, Bush D;  
XX  
DR WPI; 2003-576092/54.  
DR N-PSDB; ADA32476.  
XX

PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents  
 PT for diagnosing a bacterial disease, as components of antibacterial  
 PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for  
 PT plants.

XX Example; SEQ ID NO 7889; 328pp; English.

PS  
 XX  
 CC The invention relates to isolated Acinetobacter baumannii nucleic acids.  
 CC The A. baumannii nucleic acids and polypeptides are useful as reagents  
 CC for diagnosing a bacterial disease, as components of antibacterial  
 CC vaccines, as targets for antibacterial drugs, to detect the presence of  
 CC A. baumannii and other Acinetobacter species in a sample, in screening  
 CC compounds for the ability to interfere with the A. baumannii life cycle  
 CC or to inhibit A. baumannii infection, and as biocontrol agents for  
 CC plants. The present sequence represents the amino acid sequence of an A.  
 CC baumannii protein.

XX SQ Sequence 269 AA;

Alignment Scores: Pred. No.: 22.7 Length: 269  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservativeness: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: 6 Gaps: 0

US-09-729-264-3 (1-1168) x ADA36602 (1-269)

QY 1050 GTAGCTGTGGCCCTCCTCACAGCGGG 1076  
 DB 17 ValAlaValAlaLeuLeuThrSerGly 25

RESULT 28

AAR22653  
 ID AAR22653 standard; protein; 331 AA.

XX AC AAR22653;

XX DT 25-MAR-2003 (revised)

XX DT 09-OCT-1992 (first entry)

XX DE Transglutaminase (expressed in E. coli).

XX KW BTG; acyl rearrangement; deamination; yeast; actinomycetes.

XX OS Synthetic.

XX FN EP481504-A.

XX PD 22-APR-1992.

XX PF 18-OCT-1991; 91EP-00117813.

XX PR 19-OCT-1990; 90JP-00282566.

XX PA (AMANO) AMANO PHARM KK.

XX PA (AJIN) AJINOMOTO KK.

XX PI Takagi H, Arafuka S, Matsui H, Washizu K, Ando K, Koikeda S;

XX WPI; 1992-133808/17.

XX DR N-PSDB; AAQ24207.

XX DNA fragment encoding transglutaminase - is inserted into vector, e.g.

XX PT PhJ1053-BTG, for protein expression.

XX PS Disclosure; Page 3; 55pp; English.

XX The protein sequence given has transglutaminase (BTG) activity. The DNA  
 CC encoding this protein has a base sequence which can be used suitably in  
 CC an expression system using E. coli or yeast as a host. The base sequence  
 CC can be compared to those given in AAQ24197 and AAQ24200 which are derived

CC from Actinomycetes sp. BTG catalyses an acyl rearrangement reaction of a  
 CC gamma-carboxamide gp. of glutamine. It introduces intra- or  
 CC intermolecular formation of epsilon-(gamma-Gln)-lys cross-linking when an  
 CC epsilon-amino gp. of a Lys residue acts as an acyl receptor. When water  
 CC acts as an acyl acceptor the enzyme accelerates the conversion of Gln  
 CC residues to Glu residues by deamination. The enzyme is used in the prodn.  
 CC of gelled foods, gelled cosmetics, yogurt, gelatin, cheese etc. It is  
 CC also used in the prodn. of thermally stable materials such as  
 CC microcapsules and carriers of immobilized enzymes. The DNA sequence given  
 CC allows the prodn. of BTG efficiently and in large quantity. (Updated on  
 CC 25-MAR-2003 to correct PA field.)

XX SQ Sequence 331 AA;

Alignment Scores: Pred. No.: 22.3 Length: 331  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservativeness: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: 2 Gaps: 0

US-09-729-264-3 (1-1168) x AAR22653 (1-331)

QY 349 CCTTACCGTCCAGTTATGGAGAGCT 375

DB 19 ProTyrArgProSerTyrGlyArgAla 27

RESULT 29

AAY33665

ID AAY33665 standard; protein; 331 AA.

XX AC AAY33665;

XX DT 17-OCT-2003 (revised)

XX DT 07-JAN-2000 (first entry)

XX DE Streptovorticillum mobaraense transglutaminase protein fragment.

XX KW Transglutaminase; food industry; pharmaceutical industry; texture;  
 XX cosmetic industry; proteinaceous material; gel strength; viscosity;  
 XX breaking strength; elasticity; taste.

XX OS Streptomyces mobaraensis.

XX FN DE19814860-A1.

XX PD 07-OCT-1999.

XX PF 02-APR-1998; 98DE-01014860.

XX PR 02-APR-1998; 98DE-01014860.

XX PA (FUCH/) FUCHSBAUER H.

XX XX Fuchsbaue H, Pasternack R, Dorsch S, Otterbach J, Robenek I;

XX PI Mainusch M, Dauscher C;

XX WPI; 1999-552288/47.

XX Bacterial transglutaminase polypeptides useful for polymerizing proteins,  
 PT e.g. to modify the properties of food, pharmaceutical or cosmetic  
 PT products.

XX Claim 12; Page 25-26; 44pp; German.

XX This invention describes a novel bacterial transglutaminase polypeptide.  
 CC (I) can be used, e.g. in the food, pharmaceutical and cosmetic  
 CC industries, to polymerize proteinaceous materials in order to improve  
 CC their properties, e.g. texture, gel strength, breaking strength,  
 CC viscosity, elasticity or taste. (I) can also be used to immobilize  
 CC enzymes and antibodies. This sequence represents a transglutaminase  
 CC protein fragment isolated from Streptovorticillum mobaraense. (Updated



CC on 17-OCT-2003 to standardise OS field)

XX SQ Sequence 331 AA;

Alignment Scores:

Pred. No.:	22.3	Length:	331
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.32%	Indels:	0
DB:	2	Gaps:	0

US-09-729-264-3 (1-1168) x AAY33665 (1-331)

QY 349 CTTACCGTCCAAAGTTATGGAGAGCT 375  
 |||||  
 Db 19 ProTyArgProSerTyrglyArgAla 27

RESULT 30

AY33662

ID AAY33662 standard; protein; 331 AA.

XX AC AAY33662;

XX DT 17-OCT-2003 (revised)

XX DT 07-JAN-2000 (first entry)

XX DE Streptovorticillium S-8112 transglutaminase protein fragment.

XX KW Transglutaminase; food industry; pharmaceutical industry; texture;

XX KW cosmetic industry; proteinaceous material; gel strength; viscosity;

XX KW breaking strength; elasticity; taste.

XX OS Streptomyces sp.

XX PN DE19814860-A1.

XX PD 07-OCT-1999.

XX PF 02-APR-1998; 98DE-01014860.

XX PR 02-APR-1998; 98DE-01014860.

XX PA (FUCH/) FUCHSBAUER H.

XX PI Fuchsbaue H, Pasternack R, Dorsch S, Otterbach J, Robenek I;

XX PI Mainusch M, Dauscher C;

XX DR WPI; 1999-552288/47.

XX DR N-PSDB; AA223653.

XX PT Bacterial transglutaminase polypeptides useful for polymerizing proteins,

XX PT e.g. to modify the properties of food, pharmaceutical or cosmetic

XX PT products.

XX PS Claim 1; Page 23-24; 44pp; German.

XX CC This invention describes a novel bacterial transglutaminase polypeptide.

XX CC (1) can be used, e.g. in the food, pharmaceutical and cosmetic

XX CC industries, to polymerize proteinaceous materials in order to improve

XX CC their properties, e.g. texture, gel strength, breaking strength,

XX CC viscosity, elasticity or taste. (1) can also be used to immobilize

XX CC enzymes and antibodies. This sequence represents a transglutaminase

XX CC isolated from Streptovorticillium S-8112. (Updated on 17-OCT-2003 to

XX CC standardise OS field)

XX SQ Sequence 331 AA;

Alignment Scores:

Pred. No.:	22.3	Length:	331
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

Query Match: 2.32% Indels: 0

DB: 2 Gaps: 0

US-09-729-264-3 (1-1168) x AAY33662 (1-331)

QY 349 CTTACCGTCCAAAGTTATGGAGAGCT 375  
 |||||  
 Db 19 ProTyArgProSerTyrglyArgAla 27

RESULT 31

AAW67770

ID AAW67770 standard; protein; 331 AA.

XX AC AAW67770;

XX DT 17-OCT-2003 (revised)

XX DT 01-APR-1999 (first entry)

XX DE A transglutaminase enzyme sequence.

XX KW Transglutaminase; microbial; gelled food; jelly; yogurt; cheese;

XX KW cosmetic; meat quality; microcapsule production; high thermal stability;

XX KW carrier; immobilised enzyme.

XX OS Streptomyces sp.

XX PN EP889133-A2.

XX PD 07-JAN-1999.

XX PF 02-JUL-1998; 98EP-00112315.

XX PR 04-JUL-1997; 97JP-00180010.

XX PA (AJIN ) AJINOMOTO CO INC.

XX PI Yokoyama K, Nakamura N, Miwa T, Seguro K;

XX DR WPI; 1999-062664/06.

XX DR N-PSDB; AAV81507.

XX PT New microbial transglutaminase with N-terminal aspartic acid deleted -

XX PT allowing high level recombinant production without added methionine in E.

XX PT coli, useful in production of gelled foods, cosmetics etc.

XX PS Claim 1; Page 12-14; 56pp; English.

XX CC The present sequence represents a transglutaminase of Streptovorticillium

XX CC sp. . The specification describes a new microbial transglutaminase that

XX CC has the N-terminal aspartic acid of the present transglutaminase deleted.

XX CC Eliminating the N-terminal Asp from microbial transglutaminase allows

XX CC efficient removal of the terminal Methionine residue added when the

XX CC protein is expressed in Escherichia coli. The E. coli methionine

XX CC aminopeptidase acts well on Met-Ser but only poorly on Met-Asp, so

XX CC problems of antigenicity associated with Met-terminated proteins are

XX CC avoided. Recombinant transglutaminase is used to produce gelled foods

XX CC (jellies, yogurt and cheeses) or cosmetics, to improve the quality of

XX CC meat, in the production of materials for microcapsules of high thermal

XX CC stability and as a carrier for immobilised enzymes. (Updated on 17-OCT-

XX CC 2003 to standardise OS field)

XX SQ Sequence 331 AA;

Alignment Scores:

Pred. No.:	22.3	Length:	331
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.32%	Indels:	0
DB:	2	Gaps:	0

US-09-729-264-3 (1-1168) x AAW67770 (1-331)

QY 349 CCTTACCGTCCAGTTATGGAGAGCT 375  
 DB 19 ProTyrArgProSerTyrGlyArgAla 27

RESULT 32  
 AAB81161  
 ID AAB81161 standard; protein; 331 AA.  
 XX AAB81161;  
 XX 11-SEP-2003 (revised)  
 DT 13-JUL-2001 (first entry)  
 XX Transglutaminase related protein SEQ ID 5.  
 DE  
 XX Coryneform bacteria; transglutaminase; food processing.  
 KW  
 XX Streptomyces mobaraensis.  
 OS  
 XX WO200123591-A1.  
 PN  
 XX 05-APR-2001.  
 PD  
 XX 29-SEP-2000; 2000WO-JP006780.  
 PF  
 XX 30-SEP-1999; 99JP-00280098.  
 PR  
 XX 28-JUN-2000; 2000JP-00194043.  
 XX  
 XX (AJIN ) AJINOMOTO CO INC.  
 PA  
 XX Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Matsui H;  
 PI  
 XX WPI; 2001-266172/27.  
 DR  
 XX Efficient secretory production of foreign proteins e.g. transglutaminase  
 XX employing transformant coryneform bacterium, simply on industrial scale  
 PT with direct recovery for use in food processing and pharmaceutical  
 PT industry.  
 XX  
 XX Claim 22; Page 79-81; 151pp; Japanese.  
 PS  
 XX This invention relates to a process for the production of a foreign  
 CC secretory protein through the construction of a recombinant coryneform  
 CC bacterium. The coryneform bacterium is transformed with an expression  
 CC construct in which DNA encoding a target foreign protein pro-structure is  
 CC ligated to the downstream region of DNA encoding the signal peptide  
 CC domain of a coryneform bacterial protein. Following transformation with  
 CC the vector, the bacterium is cultured, and the pro-peptide cleaved from  
 CC the expressed protein. Transglutaminases produced using this process are  
 CC useful in the food processing and pharmaceutical industries. The present  
 CC sequence represents a transglutaminase related protein, which can be used  
 CC in the method of the invention. (Updated on 11-SEP-2003 to standardise OS  
 CC field)  
 XX  
 XX SQ Sequence 331 AA;

Alignment Scores:  
 Pred. No.: 22.3 Length: 331  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x AAB81161 (1-331)

QY 349 CCTTACCGTCCAGTTATGGAGAGCT 375  
 DB 19 ProTyrArgProSerTyrGlyArgAla 27

RESULT 33  
 ABB06742  
 ID ABB06742 standard; protein; 331 AA.

XX ABB06742;  
 AC 12-JUN-2002 (first entry)  
 DT  
 XX Streptovorticillium mobaraense transglutaminase protein SEQ ID NO:2.  
 DE  
 XX Microbial; transglutaminase; protein co-ordinate data; stereo-structure;  
 KW X-ray analysis; crystalline structure; enzyme; food processing;  
 KW thermal stability.  
 XX  
 XX Streptomyces mobaraensis.  
 OS  
 XX WO200214518-A1.  
 PN  
 XX 21-FEB-2002.  
 PD  
 XX 15-AUG-2001; 2001WO-JP007038.  
 PF  
 XX 17-AUG-2000; 2000JP-00247664.  
 PR  
 XX 27-DEC-2000; 2000JP-00396695.  
 XX  
 XX (AJIN ) AJINOMOTO CO INC.  
 PA  
 XX Kashiwagi T, Shimba N, Ishikawa K, Suzuki E, Yokoyama K;  
 XX Hirayama K;  
 PI  
 XX WPI; 2002-269198/31.  
 DR  
 XX N-PSDB; ABL50235.  
 XX  
 XX Designing and constructing a variant of Streptovorticillium mobaraense-  
 PT originated transglutaminase (MTG), for use in food processing, comprises  
 PT estimating the binding site of MTG based on its stereo-structure.  
 XX  
 XX Claim 2; Page 100-102; 126pp; Japanese.  
 PS  
 XX The present invention describes a method for designing and constructing a  
 CC variant transglutaminase by estimating the binding site of  
 CC Streptovorticillium mobaraense-originated transglutaminase (MTG) to the  
 CC substrate based on the stereo-structure obtained by X-ray analysis of the  
 CC crystalline structure of MTG crystals, and e.g. substituting amino acid  
 CC residues located at the substrate-binding site of the transglutaminase.  
 CC The method can be used for designing and constructing a variant  
 CC transglutaminase. The obtained transglutaminases can be used in food  
 CC processing. The modified transglutaminases have improved transglutaminase  
 CC activity and thermal stability, substrate-specificity and an less  
 CC required optimum pH. The present sequence represents a transglutaminase  
 CC isolated from Streptovorticillium mobaraense (also called Streptomyces  
 CC mobaraensis)  
 XX  
 XX SQ Sequence 331 AA;

Alignment Scores:  
 Pred. No.: 22.3 Length: 331  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: 5 Gaps: 0

US-09-729-264-3 (1-1168) x ABB06742 (1-331)

QY 349 CCTTACCGTCCAGTTATGGAGAGCT 375  
 DB 19 ProTyrArgProSerTyrGlyArgAla 27

RESULT 34  
 AAW67771  
 ID AAW67771 standard; protein; 332 AA.  
 XX AAW67771;  
 AC  
 XX 01-APR-1999 (first entry)  
 DT

XX DE Protein encoded by high expression transglutaminase gene.  
 XX XX  
 KW Transglutaminase; microbial; gelled food; jelly; yogurt; cheese;  
 KW cosmetic; meat quality; microcapsule production; high thermal stability;  
 KW carrier; immobilised enzyme.  
 XX XX  
 OS Synthetic.  
 OS Streptomyces sp.  
 XX XX  
 PN EP889133-A2.  
 XX XX  
 PD 07-JAN-1999.  
 XX XX  
 PF 02-JUL-1998; 98EP-00112315.  
 XX XX  
 PR 04-JUL-1997; 97JP-00180010.  
 XX XX  
 PA (AJIN ) AJINOMOTO CO INC.  
 XX XX  
 PI Yokoyama K, Nakamura N, Miwa T, Seguro K;  
 XX N-PSDB; AAV81508.  
 DR WPI; 1999-062664/06.  
 DR N-PSDB; AAV81508.  
 XX XX  
 PT New microbial transglutaminase with N-terminal aspartic acid deleted -  
 PT allowing high level recombinant production without added methionine in E.  
 PT coli, useful in production of gelled foods, cosmetics etc.  
 XX XX  
 PS Example 1; Page 18-23; 56pp; English.  
 XX XX  
 CC The present sequence is encoded by the high expression transglutaminase  
 CC gene present in plasmid pTRPMG-02. The gene is derived from  
 CC Streptococcus thermophilus sp., and is codon altered, using oligonucleotides  
 CC AAV81521-60, for expression in Escherichia coli. The specification  
 CC describes a new microbial transglutaminase that has the N-terminal  
 CC aspartic acid of transglutaminase deleted. Eliminating the N-terminal Asp  
 CC from microbial transglutaminase allows efficient removal of the terminal  
 CC Met residue added when the protein is expressed in E. coli. The E. coli  
 CC methionine aminopeptidase acts well on Met-Ser but only poorly on Met-  
 CC Asp, so problems of antigenicity associated with Met-terminated proteins  
 CC are avoided. Recombinant transglutaminase is used to produce gelled foods  
 CC (jellies, yogurt and cheeses) or cosmetics, to improve the quality of  
 CC meat, in the production of materials for microcapsules of high thermal  
 CC stability and as a carrier for immobilised enzymes  
 XX XX  
 SQ Sequence 332 AA;  
 Alignment Scores:  
 Pred. No.: 22.3 Length: 332  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: Gaps: 0  
 US-09-729-264-3 (1-1168) x AAV67771 (1-332)  
 QY 349 CCTTACCGTCCAGTTATGGGAGAGCT 375  
 Db 20 ProlyrArgProSerTyrGlyArgala 28  
 RESULT 35  
 AAB12809  
 ID AAB12809 standard; protein; 332 AA.  
 XX AAB12809;  
 AC AAB12809;  
 XX XX  
 DT 24-NOV-2000 (first entry)  
 XX XX  
 DE Transglutaminase protein sequence SEQ ID NO:1.  
 XX Transglutaminase; gelled food; jelly; yoghurt; gelled cosmetic; cheese.  
 KW

XX OS Unidentified.  
 XX OS WO200040706-A1.  
 PN WO200040706-A1.  
 XX XX  
 PD 13-JUL-2000.  
 XX XX  
 PF 24-DEC-1999; 99WO-JP007250.  
 XX XX  
 PR 28-DEC-1998; 98JP-00373131.  
 XX XX  
 PA (AJIN ) AJINOMOTO CO INC.  
 XX XX  
 PI Yokoyama K, Ono K, Ejima D;  
 XX WPI; 2000-475826/41.  
 DR N-PSDB; AAA73025.  
 XX XX  
 PT Production of active transglutaminase from denatured enzyme by two-stage  
 PT refolding process for industrial production of active enzyme for use in  
 PT food production.  
 XX XX  
 PS Disclosure; Page 48-50; 74pp; Japanese.  
 XX XX  
 CC The present invention describes a method for producing active  
 CC transglutaminase from denatured enzyme. The method comprises: (i) forming  
 CC an intermediate structure of the enzyme having transglutaminase activity  
 CC under acidic conditions in an aqueous medium; and (ii) forming a higher-  
 CC level structure of the enzyme having transglutaminase activity under  
 CC neutral conditions in an aqueous medium. The method can be used for  
 CC industrial production of active transglutaminase from denatured material  
 CC (such as recombinant transglutaminase) which can be used in the food  
 CC industry for the production of gelled foods such as jellies, yoghurts and  
 CC cheeses, and for the production of gelled cosmetics. The present sequence  
 CC represents a transglutaminase which is used in the exemplification from  
 CC the present invention  
 XX XX  
 SQ Sequence 332 AA;  
 Alignment Scores:  
 Pred. No.: 22.3 Length: 332  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: Gaps: 0  
 US-09-729-264-3 (1-1168) x AAB12809 (1-332)  
 QY 349 CCTTACCGTCCAGTTATGGGAGAGCT 375  
 Db 20 ProlyrArgProSerTyrGlyArgala 28  
 RESULT 36  
 ABB06743  
 ID ABB06743 standard; protein; 332 AA.  
 XX ABB06743;  
 AC ABB06743;  
 XX XX  
 DT 12-JUN-2002 (first entry)  
 XX XX  
 DE Streptococcus thermophilus  
 XX Microbial; transglutaminase; protein co-ordinate data; stereo-structure;  
 KW X-ray analysis; crystalline structure; enzyme; food processing;  
 KW thermal stability.  
 XX Streptomyces mobaraensis.  
 OS Streptomyces mobaraensis.  
 XX XX  
 FH Key Location/Qualifiers  
 FT Peptide 1  
 FT Protein /label= signal  
 FT 2.332

```
FT XX /label= transglutaminase
PN WO200214518-A1.
XX
XX 21-FEB-2002.
XX
XX 15-AUG-2001; 2001WO-JP007038.
XX
XX 17-AUG-2000; 2000JP-00247664.
XX
XX 27-DEC-2000; 2000JP-00396695.
XX
XX (AJIN ) AJINOMOTO CO INC.
XX
XX Kashiwagi T, Shimba N, Ishikawa K, Suzuki E, Yokoyama K;
PI Hirayama K;
XX
XX WPI; 2002-269198/31.
XX
XX N-PSDB; ABL50236.
XX
XX Designing and constructing a variant of Streptovorticillum mobaraense-
PT originated transglutaminase (MTG), for use in food processing, comprises
PT estimating the binding site of MTG based on its stereo-structure.
XX
XX Example 4; Page 105-106; 126pp; Japanese.
XX
XX The present invention describes a method for designing and constructing a
CC variant transglutaminase by estimating the binding site of
CC Streptovorticillum mobaraense-originated transglutaminase (MTG) to the
CC substrate based on the stereo-structure obtained by X-ray analysis of the
CC crystalline structure of MTG crystals, and e.g. substituting amino acid
CC residues located at the substrate-binding site of the transglutaminase.
CC The method can be used for designing and constructing a variant
CC transglutaminase. The obtained transglutaminases can be used in food
CC processing. The modified transglutaminases have improved transglutaminase
CC activity and thermal stability, substrate-specificity and an less
CC required optimum pH. The present sequence represents a transglutaminase
CC isolated from Streptovorticillum mobaraense (also called Streptomyces
CC mobaraensis)
XX
XX Sequence 332 AA;
SQ
Alignment Scores:
Pred. No.: 22.3 Length: 332
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.32% Indels: 0
DB: 5 Gaps: 0
US-09-729-264-3 (1-1168) x ABB06743 (1-332)
QY 349 CCTTACCGTCCAACTTATGGAGAGCT 375
Db ||||| 20 ProfyArgProSerTyRGlyArgala 28
RESULT 37
ADB37643
ID ADB37643 standard; protein; 332 AA.
XX
XX ADB37643;
XX
XX 04-DEC-2003 (first entry)
XX
XX Streptomyces mobaraensis transglutaminase SEQ ID NO:6.
XX
XX polyvalent antigen; pAg; antigen; Ag; crosslinking; biological agent;
XX virucide; antibacterial; fungicide; antiparasitic; cytostatic; nootropic;
XX neuroprotective; vaccine; immunogenic; antigenic; medicine; infection;
XX cancer; Alzheimer's disease; immune-related disease; autoimmune disease;
XX Streptomyces mobaraensis; transglutaminase; EC 2.3.2.13; enzyme.
XX
XX Streptomyces mobaraensis.
XX
```

```

FH Key Location/Qualifiers
FT Misc-difference 277
FT FT /note= "encoded by AT"
XX
XX WO2003074004-A2.
XX
XX 12-SEP-2003.
XX
XX 03-MAR-2003; 2003WO-US006661.
XX
XX 01-MAR-2002; 2002US-0361166P.
XX
XX 08-MAR-2002; 2002US-0363445P.
XX
XX 28-AUG-2002; 2002US-00231063.
XX
XX 28-AUG-2002; 2002US-00231114.
XX
XX 28-AUG-2002; 2002US-00231213.
XX
XX 28-AUG-2002; 2002US-00231298.
XX
XX 28-AUG-2002; 2002US-00231470.
XX
XX (CHOU/) CHOU S.
XX
XX Chou S;
XX
XX WPI; 2003-756754/71.
XX
XX N-PSDB; ADB37658.
XX
XX Preparation of polyvalent antigen, useful in vaccines, comprises
PT crosslinking antigen in presence of biological agent, especially
PT transglutaminase, and derived antibodies.
XX
XX Claim 69; Page 61-62; 130pp; English.
XX
XX The present invention describes a method for producing a polyvalent
CC antigen (pAg). The method comprises: (i) treating an antigen (Ag) in a
CC crosslinking solution with a biological agent (I); and (ii) incubating so
CC that Ag is converted to crosslinked products. The pAg has virucide,
CC antibacterial, fungicide, antiparasitic, cytostatic, nootropic and
CC neuroprotective activities, and can be used in vaccines. The method is
CC useful in preparing immunogenic compositions, using disease-specific
CC compounds optionally modified to include a (I)-reactive amino acid that
CC are combined in crosslinking solution then treated with (I), that can be
CC used as vaccines. pAg, and other related antigenic compositions, are
CC useful in human and veterinary medicine, particularly as vaccines, for
CC treatment and prevention of infections (viral, bacterial, fungal or
CC parasitic), cancers and Alzheimer's disease, also of immune-related or
CC autoimmune diseases. Antibodies against pAg are useful as diagnostic
CC reagents and crosslinked proteins can also be used industrially, e.g. in
CC food or leather processing, in cosmetics and as enzyme carriers. The
CC present sequence represents a Streptomyces mobaraensis transglutaminase
CC (EC 2.3.2.13), which is used in an example from the present invention.
XX
XX Sequence 332 AA;
SQ
Alignment Scores:
Pred. No.: 22.3 Length: 332
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.32% Indels: 0
DB: 7 Gaps: 0
US-09-729-264-3 (1-1168) x ADB37643 (1-332)
QY 349 CCTTACCGTCCAACTTATGGAGAGCT 375
Db ||||| 20 ProfyArgProSerTyRGlyArgala 28
RESULT 38
AAR49048
ID AAR49048 standard; protein; 346 AA.
XX
XX AAR49048;
XX
XX 10-MAR-2003 (revised)
DT
```

DT	20-SEP-1994	(first entry)
XX		
DE	Bacterial transglutaminase.	
XX		
KW	Bacterial transglutaminase; BTG; expression; active; inactive;	
KW	inclusion body.	
XX		
OS	Escherichia coli.	
OS	Synthetic.	
XX		
PN	JP06030771-A.	
XX		
PD	08-FEB-1994.	
XX		
PF	14-JUL-1992; 92JP-00187038.	
XX		
PR	14-JUL-1992; 92JP-00187038.	
XX		
PA	(AJIN ) AJINOMOTO KK.	
XX		
DR	WPI; 1994-079294/10.	
DR	N-PSDB; AAQ55983.	
XX		
PT	Prepn. of bacterial trans-glutaminase in large amts. - by expression of	
PT	fusion protein in E. coli bacterial trans-glutaminase.	
XX		
PS	Claim 1; Page 7-8; 13pp; Japanese.	
XX		
CC	A fused protein contains amino acids 16-346 of BTG (AAQ55983) and a	
CC	hydrophilic peptide at the amino terminal. Expression of DNA encoding	
CC	this protein in E. coli allow large scale prodn. of BTG. An active BTG	
CC	can be prepd. from the inactive fused protein inclusion body. (Updated on	
CC	10-MAR-2003 to add missing OS field.)	
XX		
SQ	Sequence 346 AA;	
	Alignment Scores:	
	Pred. No.:	22.2
	Score:	9.00
	Percent Similarity:	100.00%
	Best Local Similarity:	100.00%
	Query Match:	2.32%
	DB:	2
		Length: 346
		Matches: 9
		Conservative: 0
		Mismatches: 0
		Indels: 0
		Gaps: 0
	US-09-729-264-3 (1-1168) x AAR49048 (1-346)	
	QY 349 CTTACCGTCCCAAGTTATGGGAGAGCT 375	
	DB 34 ProTyrArgProSerTyrGlyArgala 42	
	RESULT 39	
	ADB37647	
	ID ADB37647 standard; protein, 355 AA.	
	XX	
	AC ADB37647;	
	XX	
	04-DEC-2003 (first entry)	
	XX	
	Streptomyces mobaraensis 6X-His-TGase fusion protein SEQ ID NO:10.	
	DE	
	XX	
	KW polyvalent antigen; pAg; antigen; Ag; crosslinking; biological agent;	
	KW virucide; antibacterial; fungicide; antiparasitic; cytostatic; nootropic;	
	KW neuroprotective; vaccine; immunogenic; antigenic; medicine; infection;	
	KW cancer; Alzheimer's disease; immune-related disease; autoimmune disease;	
	KW Streptomyces mobaraensis; transglutaminase; EC 2.3.2.13; enzyme;	
	KW fusion protein.	
	XX	
	OS Synthetic.	
	OS Streptomyces mobaraensis.	
	XX	
	Key Location/Qualifiers	
	PH Misc-difference 300	
	FT /note= "encoded by At"	
	PT	

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS WO200175067-A2.  
 XX 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US008631.  
 XX 31-MAR-2000; 2000US-00540217.  
 XX 23-AUG-2000; 2000US-00649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 XX N-PSDB; AAS86399.  
 DR New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX Claim 20; SEQ ID NO 52571; 103pp; English.  
 PS The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 360 AA;

Alignment Scores:  
 Pred. No.: 22.1 Length: 360  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x ABG22212 (1-360)

QY 469 CTGGAGCTCGGTCTCTCTGGTCAGCCA 495  
 Db 189 LeuGlyAlaArgSerProGlyGlnPro 197

RESULT 41  
 AAR22651  
 ID AAR22651 standard; protein; 406 AA.  
 XX  
 AC AAR22651;  
 XX

DT 27-AUG-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 09-OCT-1992 (first entry)  
 XX Transglutaminase.  
 DE  
 XX BTG; acyl rearrangement; deamination.  
 XX Fungi.  
 XX  
 XX Key Location/Qualifiers  
 FT Peptide 1..75  
 FT Protein /label= sig\_peptide  
 FT /label= mat\_BTG  
 XX  
 XX EP481504-A.  
 XX 22-APR-1992.  
 XX 18-OCT-1991; 91EP-00117813.  
 XX 19-OCT-1990; 90JP-00282566.  
 XX (AMAN ) AMANO PHARM KK.  
 XX (AJIN ) AJINOMOTO KK.  
 XX Takagi H, Arafuka S, Matsui H, Washizu K, Ando K, Koikeda S;  
 XX WPI; 1992-133808/17.  
 XX N-PSDB; AAQ24197, AAQ24201.  
 XX DNA fragment encoding transglutaminase - is inserted into vector, e.g.  
 PT PnJ1053-BTG, for protein expression.  
 XX Disclosure; Page 42; 55pp; English.  
 XX The mature transglutaminase enzyme (BTG) can be derived from two  
 CC different species, Streptovorticillium and Actinomyces. It catalyses an  
 CC acyl rearrangement reaction of a gamma-carboxyamido gp. of glutamine. It  
 CC introduces intra- or intermolecular formation of epsilon-(gamma-Gln)-Lys  
 CC cross-linking when an epsilon-amino gp. of a Lys residue acts as an acyl  
 CC receptor. When water acts as an acyl acceptor the enzyme accelerates the  
 CC conversion of Gln residues to Glu residues by deamination. The enzyme is  
 CC used in the prodn. of gelled foods, gelled cosmetics, yogurt, gelatin,  
 CC cheese etc. It is also used in the prodn. of thermally stable materials  
 CC such as microcapsules and carriers of immobilized enzymes. The DNA  
 CC sequence given allows the prodn. of BTG efficiently and in large  
 CC quantity. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 27-  
 CC AUG-2003 to correct OS field.)  
 XX SQ Sequence 406 AA;

Alignment Scores:  
 Pred. No.: 21.8 Length: 406  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: 2 Gaps: 0

US-09-729-264-3 (1-1168) x AAR22651 (1-406)

QY 349 CCTTACCGTCCAGTTATGGAGAGCT 375  
 Db 94 ProTyArgProSerTyGlyArgAla 102

RESULT 42  
 AAB97831  
 ID AAB97831 standard; protein; 407 AA.  
 XX  
 AC AAB97831;  
 XX

DT 06-AUG-2003 (revised)  
 XX 09-AUG-2001 (first entry)  
 XX  
 DE S. mobaraense IFO 13819 transglutaminase protein SEQ ID NO:4.  
 XX  
 XX Streptovorticillium cinnamomeum IFO 12852; Streptomyces; actinomycete;  
 KW Streptovorticillium mobaraense IFO 13819; transglutaminase.  
 KW  
 XX Streptomyces mobaraensis.  
 OS  
 XX WO200129187-A1.  
 EN  
 XX 26-APR-2001.  
 PD  
 XX 13-OCT-2000; 2000WO-JP007135.  
 PF  
 XX 18-OCT-1999; 99JP-00295649.  
 PR  
 XX (AJIN ) AJINOMOTO CO INC.  
 PA  
 XX Taguchi S, Momose H;  
 PI WPI; 2001-300330/31.  
 DR N-PSDB; AAH20188.  
 DR  
 XX Streptomyces sp. carrying an actinomycete-derived gene and promoter for  
 PT producing high yields of transglutaminase.  
 XX  
 XX Disclosure; Page 33-36; 41pp; Japanese.  
 PS  
 XX The present invention describes a Streptomyces sp. containing a gene  
 CC construct comprising actinomycete-derived transglutaminase gene and  
 CC promoter. Also described are methods for producing pro-transglutaminase  
 CC and active transglutaminase. The gene construct can be used in the  
 CC production of large amounts of transglutaminase. The present sequence  
 CC represents Streptovorticillium mobaraense IFO 13819 transglutaminase,  
 CC which is given in the exemplification of the present invention. (Updated  
 CC on 06-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 407 AA;  
 Alignment Scores:  
 Pred. No.: 21.8 Length: 407  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: 4 Gaps: 0  
 US-09-729-264-3 (1-1168) x AAB97831 (1-407)  
 QY 349 CCTTACCGTCCAAAGTTATGGGAGGCT 375  
 DB 95 ProTyArgProSerTyArgGala 103  
 RESULT 43  
 AAB81166  
 ID AAB81166 standard; protein; 407 AA.  
 XX  
 AC AAB81166;  
 XX  
 XX 11-SEP-2003 (revised)  
 DT 13-JUL-2001 (first entry)  
 DT  
 XX Prepro-transglutaminase amino acid sequence.  
 DE  
 XX Coryneform bacteria; transglutaminase; food processing.  
 KW Streptomyces mobaraensis.  
 XX  
 OS WO200123591-A1.  
 PN  
 XX 05-APR-2001.  
 PD

XX 29-SEP-2000; 2000WO-JP006780.  
 PF  
 XX 30-SEP-1999; 99JP-00280098.  
 PR  
 XX 28-JUN-2000; 2000JP-00194043.  
 PR  
 XX (AJIN ) AJINOMOTO CO INC.  
 PA  
 XX Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Matsui H;  
 PI WPI; 2001-266172/27.  
 DR N-PSDB; AAF86283.  
 DR  
 XX Efficient secretory production of foreign proteins e.g. transglutaminase  
 PT employing transformant coryneform bacterium, simply on industrial scale  
 PT with direct recovery for use in food processing and pharmaceutical  
 PT industry.  
 XX  
 XX Example 1; Page 90-93; 151pp; Japanese.  
 PS  
 XX This invention relates to a process for the production of a foreign  
 CC secretory protein through the construction of a recombinant coryneform  
 CC bacterium. The coryneform bacterium is transformed with an expression  
 CC construct in which DNA encoding a target foreign protein pro-structure is  
 CC ligated to the downstream region of DNA encoding the signal peptide  
 CC domain of a coryneform bacterial protein. Following transformation with  
 CC the vector, the bacterium is cultured, and the pro-peptide cleaved from  
 CC the expressed protein. Transglutaminases produced using this process are  
 CC useful in the food processing and pharmaceutical industries. The present  
 CC sequence prepro-transglutaminase. The protein is used in an example  
 CC illustrating the method of the invention. (Updated on 11-SEP-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 407 AA;  
 Alignment Scores:  
 Pred. No.: 21.8 Length: 407  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: 4 Gaps: 0  
 US-09-729-264-3 (1-1168) x AAB81166 (1-407)  
 QY 349 CCTTACCGTCCAAAGTTATGGGAGGCT 375  
 DB 95 ProTyArgProSerTyArgGala 103  
 RESULT 44  
 ABU07390  
 ID ABU07390 standard; protein; 407 AA.  
 XX  
 AC ABU07390;  
 XX  
 XX 23-OCT-2003 (revised)  
 DT 28-JAN-2003 (first entry)  
 DT  
 XX Foreign protein #1.  
 DE  
 XX Coryneform bacterium; signal peptide domain; food processing; medicine;  
 KW cosmetic; transglutaminase; human epithelial growth factor.  
 XX  
 OS Streptomyces mobaraensis.  
 XX  
 XX WO200281694-A1.  
 PN  
 XX 17-OCT-2002.  
 PD  
 XX 27-MAR-2002; 2002WO-JP002978.  
 PF  
 XX 30-MAR-2001; 2001JP-00098808.  
 PR  
 XX



PA (AJIN ) AJINOMOTO CO INC.  
XX  
PI Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Heima H, Matsui H;  
XX  
DR WPI; 2003-058550/05.  
DR N-PSDB; ABX10285.  
XX  
XX Secretion production of foreign proteins by culturing transformant  
PT coryneform bacteria, applicable in producing e.g. industrially-useful  
PT transglutaminase and human epithelial growth factor.  
XX  
XX Disclosure; Page 68-71; 117pp; Japanese.  
XX  
XX The invention relates to a method for producing a foreign protein by  
CC culturing a mutant strain of a coryneform bacterium that contains an  
CC expression gene construct to secrete the foreign protein, followed by  
CC recovery of the produced foreign protein. The gene construct is obtained  
CC by ligation of a nucleic acid encoding a signal peptide domain  
CC originating from a coryneform bacterium, to downstream of a functioning  
CC promoter sequence in the coryneform bacterium and also by ligation of a  
CC nucleic acid encoding a foreign protein, to downstream of a nucleic acid  
CC sequence encoding the signal peptide. The method is useful for the  
CC production of a foreign protein which is applicable in producing e.g.  
CC industrially-useful transglutaminase and human epithelial growth factor  
CC for use in medicine, cosmetics and food processing. This sequence  
CC represents a foreign protein of the invention. (Updated on 23-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 407 AA;  
  
Alignment Scores:  
Pred. No.: 21.8 Length: 407  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.32% Indels: 0  
DB: 6 Gaps: 0  
  
US-09-729-264-3 (1-1168) x ABU07390 (1-407)  
  
Qy 349 CCTTACCGTCCAAGTTATGGAGAGCT 375  
Db 95 ProTyrArgProSerTyrGlyArgAla 103  
  
RESULT 45  
AAR49049  
ID AAR49049 standard; protein; 408 AA.  
XX AC AAR49049;  
XX  
XX 10-MAR-2003 (revised)  
DT 20-SEP-1994 (first entry)  
XX  
XX BTG-contg. sequence.  
DE  
XX  
XX Bacterial transglutaminase; BTG; expression; active; inactive;  
KW inclusion body.  
XX  
XX Escherichia coli.  
OS Synthetic.  
XX  
XX JP06030771-A.  
FN  
XX  
XX 08-FEB-1994.  
PD  
XX  
XX 14-JUL-1992; 92JP-00187038.  
PF  
XX  
XX 14-JUL-1992; 92JP-00187038.  
PR  
XX  
XX (AJIN ) AJINOMOTO KK.  
PA  
XX  
XX WPI; 1994-079294/10.  
DR N-PSDB; AAQ55984.  
DR

XX Prepn. of bacterial trans-glutaminase in large amts. - by expression of  
PT fusion protein in E. coli bacterial trans-glutaminase.  
XX  
XX Disclosure; Page 8-10; 13pp; Japanese.  
PS  
XX A fused protein contains amino acids 16-346 of BTG (AAQ55983) and a  
CC hydrophilic peptide at the amino terminal. Expression of DNA encoding  
CC this protein in E. coli allow large scale prodn. of BTG. An active BTG  
CC can be prepd. from the inactive fused protein inclusion body. (Updated on  
CC 10-MAR-2003 to add missing OS field.)  
XX  
SQ Sequence 408 AA;  
  
Alignment Scores:  
Pred. No.: 21.8 Length: 408  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.32% Indels: 0  
DB: 2 Gaps: 0  
  
US-09-729-264-3 (1-1168) x AAR49049 (1-408)  
  
Qy 349 CCTTACCGTCCAAGTTATGGAGAGCT 375  
Db 96 ProTyrArgProSerTyrGlyArgAla 104  
  
Search completed: September 18, 2004, 22:53:20  
Job time : 67.0692 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 18, 2004, 22:37:26 ; Search time 56.4053 Seconds

(without alignments)  
11771.718 Million cell updates/sec

Title: US-09-729-264-1

Perfect score: 389

Sequence: 1 ctgtctgccatctgaataa.....gtaatacaactgtagtatag 1175

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2094264

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Listing first 135 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/qsn2\_1/USPTO.spool\_p/US07929264/runat\_17092004\_155106\_2337/app\_query.fasta\_1.4117  
-DB=A Geneseq\_29Jan04 -QFWT=fastan -SUFFIX=oli.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=135 -DOCALIGN=200 -THR SCORE=quality -THR MIN=0 -ALIGN=45 -MODE=LOCAL  
-OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09729264 @CGN 1 1 321 @runat\_17092004\_155106\_2337 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THRSADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : A Geneseq\_29Jan04:\*

1: geneseqp1980s: \*  
2: geneseqp1990s: \*  
3: geneseqp2000s: \*  
4: geneseqp2001s: \*  
5: geneseqp2002s: \*  
6: geneseqp2003as: \*  
7: geneseqp2003bs: \*  
8: geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	382	98.2	382	5	Aau75540 Human B7-
2	316	81.2	386	5	Aau75542 Human B7-
3	316	81.2	407	7	AdB64920 Human pro
4	257	66.1	386	5	Aau75541 Human B7-
5	243	62.5	463	4	ABG28169 Novel hum
6	208	53.5	377	5	Aau75543 Human B7-
C 7	12	3.2	78	4	Aam20679 Peptide #
C 8	12	3.2	78	4	ABb42042 Peptide #
C 9	12	3.2	78	4	Aam35844 Peptide #
C 10	12	3.2	78	4	ABb25654 Protein #

C 11	12	3.2	78	4	AAM75735	Aam75735 Human bon
C 12	12	3.2	78	4	AAM62923	Aam62923 Human bra
C 13	12	3.2	78	4	ABG57473	Abg57473 Human liv
C 14	12	3.2	78	5	ABG45219	Abg45219 Human pep
C 15	11	2.9	524	4	AAU07370	Aau07370 G protein
C 16	11	2.8	1805	4	ABB65262	Abb65262 Drosophil
C 17	10	2.6	85	3	AAV65872	Aav65872 n-myc mut
C 18	10	2.6	94	3	AAV65874	Aav65874 n-myc mut
C 19	10	2.6	95	3	AAV65875	Aav65875 n-myc mut
C 20	10	2.6	2168	4	ABB64563	Abb64563 Drosophil
C 21	9	2.4	96	4	ABB69039	Abb69039 Drosophil
C 22	9	2.4	191	4	ABB63572	Abb63572 Drosophil
C 23	9	2.4	191	4	ABB70426	Abb70426 Drosophil
C 24	9	2.4	191	4	ABB67144	Abb67144 Drosophil
C 25	9	2.3	253	6	ADA48420	Ada48420 Rice prot
C 26	9	2.3	258	6	ABU17237	Abu17237 Protein e
C 27	9	2.3	269	6	ADA36602	Ada36602 Acinetoba
C 28	9	2.3	331	2	AAR22653	Aar22653 Transglut
C 29	9	2.3	331	2	AAV33665	Aav33665 Streptove
C 30	9	2.3	331	2	AAV33662	Aav33662 Streptove
C 31	9	2.3	331	2	AAW67770	Aaw67770 A transgl
C 32	9	2.3	331	4	ABb81161	Abb81161 Transglut
C 33	9	2.3	331	5	ABb06742	Abb06742 Streptove
C 34	9	2.3	332	2	AAM67771	Aam67771 Protein e
C 35	9	2.3	332	3	AAb12809	Aab12809 Transglut
C 36	9	2.3	332	5	ABb06743	Abb06743 Streptove
C 37	9	2.3	332	7	ADB37643	Adb37643 Streptomy
C 38	9	2.3	346	2	AAR49048	Aar49048 Bacterial
C 39	9	2.3	355	7	ADB37647	Adb37647 Streptomy
C 40	9	2.3	360	4	ABG22212	Abg22212 Novel hum
C 41	9	2.3	406	2	AAR22651	Aar22651 Transglut
C 42	9	2.3	407	4	AAb97831	Aab97831 S. mobara
C 43	9	2.3	407	4	ABb81166	Abb81166 Prepro-tr
C 44	9	2.3	407	6	ABU07390	Abu07390 Foreign p
C 45	9	2.3	408	2	AAR49049	Aar49049 BTG-contg
C 46	9	2.4	494	4	ABb68018	Abb68018 Drosophil
C 47	9	2.4	567	4	ABb65901	Abb65901 Drosophil
C 48	9	2.4	567	4	ABb64078	Abb64078 Drosophil
C 49	9	2.4	567	4	ABb66007	Abb66007 Drosophil
C 50	9	2.3	820	7	ADe65531	Ade65531 Streptove
C 51	9	2.4	904	4	ABb71477	Abb71477 Drosophil
C 52	9	2.4	904	5	ABG70014	Abg70014 Larval vi
C 53	9	2.4	908	6	ABU08492	Abu08492 Alpha-hel
C 54	9	2.4	1076	6	ABU81145	Abu81145 Human PRO
C 55	9	2.4	1076	6	ABU66845	Abu66845 Human PRO
C 56	9	2.4	1301	4	ABb65129	Abb65129 Drosophil
C 57	9	2.4	1379	4	ABb68940	Abb68940 Drosophil
C 58	9	2.4	1601	4	ABb60248	Abb60248 Drosophil
C 59	9	2.4	3972	3	AAb23749	Aab23749 S. averni
C 60	9	2.4	3972	4	AAG65264	Aag65264 Streptomy
C 61	9	2.4	3972	4	AAG65268	Aag65268 Streptomy
C 62	9	2.4	4440	6	ABU88256	Abu88256 Novel hum
C 63	9	2.4	4440	6	ABU90135	Abu90135 Novel hum
C 64	9	2.4	4440	6	ABU96437	Abu96437 Novel hum
C 65	9	2.4	4440	6	ABU99046	Abu99046 Novel hum
C 66	9	2.4	4440	6	ABU98261	Abu98261 Novel hum
C 67	9	2.4	4440	6	ABU91967	Abu91967 Novel hum
C 68	9	2.4	4440	6	ABU85271	Abu85271 Novel hum
C 69	9	2.4	4440	6	ABO00410	Abu00410 Novel hum
C 70	9	2.4	4440	6	ABU88961	Abu88961 Novel hum
C 71	9	2.4	4440	6	ABO06457	Abu06457 Novel hum
C 72	9	2.4	4440	6	ABU95517	Abu95517 Novel hum
C 73	9	2.4	4440	6	ABU95207	Abu95207 Novel hum
C 74	9	2.4	4440	6	ABU90755	Abu90755 Novel hum
C 75	9	2.4	4440	6	ABU93917	Abu93917 Novel hum
C 76	9	2.4	4440	6	ABU86191	Abu86191 Novel hum
C 77	9	2.4	4440	6	ABU82046	Abu82046 Novel hum
C 78	9	2.4	4440	6	ABU07907	Abu07907 Novel hum
C 79	9	2.4	4440	6	ABU94227	Abu94227 Novel hum
C 80	9	2.4	4440	6	ABO00100	Abu00100 Novel hum
C 81	9	2.4	4440	6	ABU87111	Abu87111 Novel hum
C 82	9	2.4	4440	6	ABU91352	Abu91352 Novel hum
C 83	9	2.4	4440	6	ABU90445	Abu90445 Novel hum

C 84 9 2.4 4440 6 ABU97036 Novel hum  
C 85 9 2.4 4440 6 ABO5232 Novel hum  
C 86 9 2.4 5532 3 Aab23752 S. avermi  
C 87 9 2.4 5532 4 Aag65267 Streptomy  
C 88 9 2.4 6025 4 Aag84939 Shrimp wh  
C 89 9 2.4 6077 6 Abp95996 White spo  
C 90 8 2.1 36 4 Aab15591 Human ner  
C 91 8 2.1 48 2 Aay12003 Human 5'  
C 92 8 2.1 50 4 Aam21199 Peptide #  
C 93 8 2.1 50 4 Aab43522 Peptide #  
C 94 8 2.1 50 4 Aam37416 Peptide #  
C 95 8 2.1 50 4 Aab26481 Protein #  
C 96 8 2.1 50 4 Aam77263 Human bon  
C 97 8 2.1 50 4 Aam64459 Human bra  
C 98 8 2.1 50 4 Aag58900 Human liv  
C 99 8 2.1 50 5 Aag46280 Human pep  
C 100 8 2.1 79 5 Abp05686 Human ORF  
C 101 8 2.1 81 4 Aab70231 Drosophil  
C 102 8 2.1 82 4 Aay76046 Murine sk  
C 103 8 2.1 82 4 Aab55985 Skin cell  
C 104 8 2.1 82 5 Aab72185 Murine pr  
C 105 8 2.1 89 4 Aam18746 Peptide #  
C 106 8 2.1 89 4 Aab37844 Peptide #  
C 107 8 2.1 89 4 Aam31246 Peptide #  
C 108 8 2.1 89 4 Aab23106 Protein #  
C 109 8 2.1 89 4 Aam70960 Human bon  
C 110 8 2.1 89 4 Aam58466 Human bra  
C 111 8 2.1 89 4 Aag52681 Human liv  
C 112 8 2.1 89 5 Aag40758 Human pep  
C 113 8 2.1 89 5 Abp32378 Human red  
C 114 8 2.1 100 3 Aab40618 Human ORF  
C 115 8 2.1 100 5 Aab02711 Human ORF  
C 116 8 2.1 106 3 Aab42377 Human ORF  
C 117 8 2.1 119 4 Aab69357 Drosophil  
C 118 8 2.1 134 4 Aab63906 Drosophil  
C 119 8 2.1 140 4 Aab71305 Drosophil  
C 120 8 2.1 148 5 AAU96222 Human sec  
C 121 8 2.1 148 5 Aag64793 Human alb  
C 122 8 2.1 149 4 Aab69455 Drosophil  
C 123 8 2.1 150 4 Aab43063 Propionib  
C 124 8 2.1 150 5 Aag61831 Prostata  
C 125 8 2.1 150 6 Aam39582 Propionib  
C 126 8 2.1 153 7 ADC08113 Rice prot  
C 127 8 2.1 155 4 Aab68931 Neisseria  
C 128 8 2.1 165 4 Aag29283 Novel hum  
C 129 8 2.1 181 4 Aag28127 Novel hum  
C 130 8 2.1 184 4 Aab66043 Drosophil  
C 131 8 2.1 185 4 Aab95364 Human pro  
C 132 8 2.1 220 4 AAU51099 Propionib  
C 133 8 2.1 220 6 Aam47618 Propionib  
C 134 8 2.1 223 5 AAU75546 Mouse B7-  
C 135 8 2.1 224 6 ADA48140 Rice prot

## ALIGNMENTS

RESULT 1  
AAU75540  
ID AAU75540 standard; protein; 382 AA.

AC AAU75540;

DT 23-APR-2002 (first entry)

DE Human B7-like protein, B7-L\_h1.

XX Human; B7-like protein; B7-L; antiinfertility; gynaecological;  
KW antitumour; cytostatic; immunosuppressive; antiarthritic; antineumatic;  
KW antiinflammatory; dermatological; antipsoriatic; neuroprotective;  
KW antidiabetic; haemostatic; antithyroid; antitumor; antiallergic;  
KW antiasthmatic; nephroprotective; antibacterial; virucide; tumour; cancer;  
KW reproductive disorder; graft versus host disease; autoimmune disease;  
KW toxic shock syndrome; allergy; nephropathy; skin disorder;

KW endocrinopathy; lymphoproliferative disorder.

XX Homo sapiens.

XX WO200200710-A2.

XX 03-JAN-2002.

XX 28-JUN-2001; 2001WO-US020719.

XX 28-JUN-2000; 2000US-0214512P.

XX 28-NOV-2000; 2000US-00729264.

(AMGE-) AMGEN INC.

Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;

WPI; 2002-130881/17.  
N-PSDB; ABK13028.

New B7-like polypeptides, polynucleotides and their modulators, useful for diagnosing, preventing and treating reproductive, immune and proliferative disorders, e.g. cancer and arteriosclerosis.

Claim 13; Fig 1; 135pp; English.

The invention relates to an isolated B7-like (B7-L) polypeptide (I). The polypeptide, polynucleotide encoding it and antibody against (I) are useful for treating B7-like polypeptide-related disease, disorders or conditions including reproductive disorders (e.g. infertility, miscarriage, preterm labour and delivery and endometriosis) and proliferative disorders. Antibodies, soluble proteins comprising extracellular domains and other regulators of B7-L polypeptides are useful for enhancing the immune response to tumours. (I) plays a role in growth and maintenance of cancer cells based on the observation of seminal vesicle hyperplasia in transgenic mice overexpressing B7-L polypeptide. Hence modulators of (I) are useful for the treatment of cancer including seminal vesicle cancer, lung, brain, breast, ovarian, testicular cancer and cancers of haematopoietic system. B7-L polypeptide pathway can be manipulated to regulate cytotoxic T-lymphocyte response in allograft transplantation, graft versus host disease, T-cell dependent B-cell mediated diseases and autoimmune diseases. B7-L molecules are useful for alleviating the symptoms associated with diseases involving chronic immune cell dysfunction or to treat autoimmune diseases such as systemic lupus erythematosus, rheumatoid arthritis, multiple sclerosis, diabetes, immune thrombocytopenic purpura and psoriasis, chronic inflammatory disease such as inflammatory bowel disease (Crohn's disease and ulcerative colitis), Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They are also useful as immunosuppressive agents for bone marrow and organ transplantation or to prolong graft survival. B7-L molecules are also useful for diagnosis and treatment of diseases involving abnormal cell proliferation, including arteriosclerosis and vascular restenosis. Antagonists of B7-L polypeptides are useful for alleviation of toxic shock syndrome or allosensitisation due to blood transfusions, and for treatment of allergy, asthma and hypersensitivity reactions, nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and pemphigoid), endocrinopathies (Grave's disease), various pneumopathies (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia, thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and lymphoproliferative disorders such as multiple myeloma. The present sequence represents the amino acid sequence of human B7-L\_h1

SQ Sequence 382 AA;

Alignment Scores:

Pred. No.:	0	Length:	382
Score:	382.00	Matches:	382
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	98.20%	Indels:	0
DB:	5	Gaps:	0

US-09-729-264-1 (1-1175) x AAU75540 (1-382)

QY	27	ATGGGCTTTGATTTTCTCCACGGTCTCTGGTCTGTGTAATGAAGTCATAGAAGCCCC	86
Db	1	MetGlyLeuValIlePheLeuHisGlySerGlySerGlyAsnGluValIleGluGlyPro	20
QY	87	CAGAAATGCAACAGTCTGAAGGCTCCACGGCTCCAGCTCGCTTCACTGCACCTCTCCACAGGC	146
Db	21	GlnAsnAlaThrValLeuLysGlySerGlnAlaArgPheAsnCysThrValSerGlnGly	40
QY	147	TGGAAGCTCATATGTCGGCTCTCAGTGACATGGTGTGCTAAGCTCAGGCCCATGGAG	206
Db	41	TrpLysLeuIleMetTrpAlaLeuSerAspMetValValLeuSerValArgProMetGlu	60
QY	207	CCATCATCAACCAATGACCGCTTACCTCTCAGAGGTACGACCGCGGGAACCTTCACC	266
Db	61	ProIleThrAsnAspArgPheThrSerGlnArgTyrAspGlnGlyGlyAsnPheThr	80
QY	267	TCGGAGATGATCATCCACATGTCGAGCCCGAGTTCGGGGAACATCAGATGCAGCCCTC	326
Db	81	SerGluMetIleIleHisAsnValGluProSerAspSerGlyAsnIleArgCysSerLeu	100
QY	327	CAGAACAGTCGCTGCATCGATCTGCTTACCTTACCGTCCCAAGTTATGGAGAGCTGTTC	386
Db	101	GlnAsnSerArgLeuHisGlySerAlaLysLeuThrValGlnValMetGlyGluLeuPhe	120
QY	387	ATTCCAGTGTAAATCTTGTACTGCTGAGATGAACCTTGTCAAGTTACTTGTCTACCC	446
Db	121	IleProSerValAsnLeuValValAlaGluAsnGluProCysGluValThrCysLeuPro	140
QY	447	TCACACTGGACCGCTCCGGATATTTCTCTGGAGCTGGTCTCTGTGTCAGCCATTC	506
Db	141	SerHisTrpThrArgLeuProAspIleSerTrpGluLeuGlyLeuValSerHisSer	160
QY	507	AGCTATTATTTGTTCCGGAGCCCGACCTTCAAAGTCAGTGCATCTCGTCTGCTG	566
Db	161	SerTyrTyrPheValProGluProSerAspLeuGlnSerAlaValSerIleLeuAlaLeu	180
QY	567	ACCCACAGAGCAATGGGACTTTGACTTCGCTGGCTACCTGGAAGAGCCTGAAGCCCGC	626
Db	181	ThrProGlnSerAsnGlyThrLeuThrCysValAlaThrTyrLysSerLeuLysAlaArg	200
QY	627	AGTCTGCACTGTAAATCTCACTGATTCGGTGTCCCAAGACACTGGAGGTGTATT	686
Db	201	LysSerAlaThrValAsnLeuThrValIleArgCysProGlnAspThrGlyGlyIle	220
QY	687	AATATTCAGGTGATTATCAAGTTTACCGAGTTTAGGTTTTCATTGCTACTTGGGCG	746
Db	221	AsnIleProGlyValLeuSerSerLeuProSerLeuGlyPheSerLeuProThrTrpGly	240
QY	747	AAAGTTGGACTTGGACTAGCAGGCCATCGCTTCTGACCGCGAGTGTACTTTACAATA	806
Db	241	LysValGlyLeuGlyLeuAlaGlyThrMetLeuLeuThrProThrCysThrLeuThrIle	260
QY	807	CGCTCTGCTGCTCGCGCTGTTGTGTGCTGCTCACTGCTGCTGCTGCTGCTGCTGCT	866
Db	261	ArgCysCysCysArgArgArgCysCysGlyCysAsnCysCysCysArgCysCysPhe	280
QY	867	TGCTGTAGAAAAAGAGGATTTCTGTTTCAATTTCAAAAGAAATCTGAAAAAGAGAG	926
Db	281	CysCysArgGlyArgGlyPheArgIleGlnPheGlnLysLysSerGluLysGluLys	300
QY	927	ACAAACAAAGAACTGAGACAGAAAGTGAATGAAATCCCGGCTTACAAATTCAGATGA	986
Db	301	ThrAsnLysGluThrGluThrGluSerGlyAsnGluAsnSerGlyTyrAsnSerAspGlu	320
QY	987	CAAAGACACAGACCGCTTCTCTCCCTCCAAATCTGTGAATCCAGTCAATCTGAA	1046
Db	321	GlnLysThrThrAspThrAlaSerLeuProProLysSerCysGluSerSerAspProGlu	340
QY	1047	CAAGAAACAGTAGCTGGCGCTCTCTACACAGCGGCTGATCAAGCTCCACCCAGGCCA	1106
Db	341	GlnArgAsnSerCysGlyProProHisGlnArgAlaAspGlnArgProProArgPro	360
QY	1107	GCAAGTATCCACAGGCTTCTTTTAACTGCGCAGTCTCGAAGGTCTGATATACT	1166
Db	361	AlaSerHisProGlnAlaSerPheAsnLeuAlaSerProGluLysValSerAsnThrThr	380
QY	1167	GTAGTA 1172	
Db	381	ValVal 382	
RESULT 2			
ID	AAU75542	standard; protein; 386 AA.	
AC	AAU75542;		
DT	23-APR-2002	(first entry)	
DE	Human B7-like protein, B7-L_h3.		
KW	Human; B7-like protein; B7-L; antiinfertility; gynaecological;		
KW	antitumour; cytostatic; immunosuppressive; antiarthritic; anti-		
KW	antiinflammatory; dermatological; antipsoriatic; neuroprotective;		
KW	antiasthmatic; haemostatic; antithyroid; antitumor; antiallergic;		
KW	reproductive disorder; antibacterial; virucide; tumour; cancer;		
KW	toxic shock syndrome; allergy; nephropathy; skin disorder;		
KW	endocrinopathy; lymphoproliferative disorder.		
OS	Homo sapiens.		
XX	WO200200710-A2.		
XX	03-JAN-2002.		
XX	28-JUN-2001; 2001WO-US020719.		
XX	28-JUN-2000; 2000US-0214512P.		
PR	28-NOV-2000; 2000US-00729264.		
XX	(AMGE-) AMGEN INC.		
PI	Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;		
XX	WPI; 2002-130881/17.		
DR	N-PSDB; ABK13030.		
XX	New B7-like polypeptides, polynucleotides and their modulators, useful		
PT	for diagnosing, preventing and treating reproductive, immune and		
PT	proliferative disorders, e.g. cancer and arteriosclerosis.		
XX	Claim 13; Fig 3; 135pp; English.		
CC	The invention relates to an isolated B7-like (B7-L) polypeptide (I). The		
CC	polypeptide, polynucleotide encoding it and antibody against (I) are		
CC	useful for treating B7-like polypeptide-related disease, disorders or		
CC	conditions including reproductive disorders (e.g. infertility,		
CC	miscarriage, preterm labour and delivery and endometriosis) and		
CC	proliferative disorders. Antibodies, soluble proteins comprising		
CC	extracellular domains and other regulators of B7-L polypeptides are		
CC	useful for enhancing the immune response to tumours. (I) plays a role in		
CC	growth and maintenance of cancer cells based on the observation of		
CC	seminal vesicle hyperplasia in transgenic mice overexpressing B7-L		
CC	polypeptide. Hence modulators of (I) are useful for the treatment of		
CC	cancer including seminal vesicle cancer, lung, brain, breast, ovarian,		
CC	testicular cancer and cancers of haematopoietic system. B7-L polypeptide		
CC	pathway can be manipulated to regulate cytotoxic T-lymphocyte response in		
CC	allograft transplantation, graft versus host disease, T-cell dependent B-		
CC	cell mediated diseases and autoimmune diseases. B7-L molecules are useful		
CC	for alleviating the symptoms associated with diseases involving chronic		
CC	immune cell dysfunction or to treat autoimmune diseases such as systemic		
CC	lupus erythematosus, rheumatoid arthritis, multiple sclerosis, diabetes,		
CC	immune thrombocytopenic purpura and psoriasis, chronic inflammatory		
CC	disease such as inflammatory bowel disease (Crohn's disease and		
CC	ulcerative colitis), Grave's disease, Hashimoto's thyroiditis and		

CC diabetes mellitus. They are also useful as immunosuppressive agents for  
 CC bone marrow and organ transplantation or to prolong graft survival. B7-L  
 CC molecules are also useful for diagnosis and treatment of diseases  
 CC involving abnormal cell proliferation, including arteriosclerosis and  
 CC vascular restenosis. Antagonists of B7-L polypeptides are useful for  
 CC alleviation of toxic shock syndrome or allo sensitisation due to blood  
 CC transfusions, and for treatment of allergy, asthma and hypersensitivity  
 CC reactions, nephropathies (e.g. glomerulonephritis), skin disorders  
 CC (pemphigus and pemphigoid), endocrinopathies (Grave's disease), various  
 CC pneumopathies (extrinsic alveolitis), vasculopathies, coeliac disease,  
 CC anaemia, thrombocytopenias, Guillain-Barre syndrome and myasthenia  
 CC gravis, and lymphoproliferative disorders such as multiple myeloma. The  
 CC present sequence represents the amino acid sequence of human B7-L<sub>h3</sub>  
 XX  
 SQ Sequence 386 AA;

Alignment Scores:  
 Pred. No.: 1,82e-296 Length: 386  
 Score: 316.00 Matches: 316  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 81.23% Indels: 0  
 DB: 5 Gaps: 0

US-09-729-264-1 (1-1175) x AAU75542 (1-386)

QY 51 GGTTCCTGGTCTGTAATGAAGTCAATAGAGCGCCCAAGATGCAACAGCTCTGAAGGC 110  
 Db 13 GlySerGlySerGlyAsnGluValIleGluGlyProGlnAsnAlaThrValLeuLysGly 32  
 QY 111 TCCAGGCTCGTTCACATGACCGTCTCCAGGGCTGGAAGCTCATCATGTGGCTCTC 170  
 Db 33 SerGlnAlaArgPheAsnCysThrValSerGlnGlyTrpLysLeuIleMetTrpAlaLeu 52  
 QY 171 AGTACATGTGTCTAAGCGTCAGCGCCATGAGCCCATCATCCAAATGACCGCTTC 230  
 Db 53 SerAspMetValValLeuSerValArgProMetGluProIleIleThrAsnAspArgPhe 72  
 QY 231 ACCTCTCAGAGGTACAGCCAGCGGGGAACTTCACCTCGGAGATGATCATCCAAATGTG 290  
 Db 73 ThrSerGlnArgTrpAspGlnGlyGlyAsnPheThrSerGluMetIleIleHisAsnVal 92  
 QY 291 GAGCCCATGATTCGGGGAACATCAGATGACGCTCCAGAGCTCCAGAACATGCGCTCATGATCT 350  
 Db 93 GluProSerAspSerGlyAsnIleArgCysSerLeuGlnAsnSerArgLeuHisGlySer 112  
 QY 351 GCTTACCTTACCGTCCCAAGTTATGGAGAGCTGTTCATCCAGTGTAAATCTGTAGTC 410  
 Db 113 AlaTrpLeuThrValGlnValMetGlyGluLeuPheIleProSerValAsnLeuVal 132  
 QY 411 GCTGAGAAATGAACCTTGTGAAGTTACTTGTCTACCTCAGAGTGGAGCCCGCTCCCGAT 470  
 Db 133 AlaGluAsnGluProCysGluValThrCysLeuProSerHisTrpThrArgLeuProAsp 152  
 QY 471 ATTTCCTGGAGCTCGGTCTCTCGTCAGCCATTCAGCTATTATTGTTCGGAGCCC 530  
 Db 153 IleSerTrpGluLeuGlyLeuLeuValSerHisSerTrpTrpPheValProGluPro 172  
 QY 531 AGCGACTTCAAAGTGCAGTGAGCATCTGGCTCTGACCCACAGAGCAATGGGACTTGT 590  
 Db 173 SerAspLeuGlnSerAlaValSerIleLeuAlaLeuThrProGlnSerAsnGlyThrLeu 192  
 QY 591 ACTTGGCTGGCTACCTGGAAGAGCTGAAGCCCGCAAGTCTGCAACTGTAAATCTCACT 650  
 Db 193 ThrCysValAlaThrTrpLysSerLeuLysAlaArgLysSerAlaThrValAsnLeuThr 212  
 QY 651 GTGATTCCGCTCCCGACACACGAGTGGAGTGTATTAATATTCAGGTGTTATTCAGT 710  
 Db 213 ValIleArgCysProGlnAspThrGlyGlyIleAsnIleProGlyValLeuSerSer 232  
 QY 711 TTACCGAGTTTATAGTTTTCATTGCTACTTGGGGCAAGTTGGACTTGGACTAGCAGGC 770  
 Db 233 LeuProSerLeuGlyPheSerLeuProThrTrpGlyLysValGlyLeuGlyLeuAlaGly 252

QY 771 ACCATGCTTCTGACGGCAGCTGCTACTTCAATACGCTGCTCTCTCCGCCGCTGT 830  
 Db 253 ThrMetLeuLeuThrProThrCysThrLeuThrIleArgCysCysCysArgArgArg 272  
 QY 831 TGTTGTGGCTGCAACTGCTCTGCTCGGTGTGTTGTTCTGCTGTAGAGAAAAGGATTT 890  
 Db 273 CysCysGlyCysAsnCysCysCysArgCysCysPheCysCysArgLysArgGlyPhe 292  
 QY 891 CGTATTCAATTTCAAAGAAATCTGAAAAAGAGAGACAAACAAAGAACTGACACAGAA 950  
 Db 293 ArgIleGlnPheGlnLysLysSerGluLysGluLysThrAsnLysGluThrGluThr 312  
 QY 951 AGTGGAAATGAAATCTCCGGCTCAATTCAGATGCAACAAAAGACCACA 998  
 Db 313 SerGlyAsnGluAsnSerGlyTyrAsnSerAspGluGlnLysThrThr 328

# RESULT 3

ADB64920  
 ID ADB64920 standard; protein; 407 AA.  
 XX  
 AC ADB64920;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human protein encoded by clone PLACB60177880.  
 XX  
 KW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;  
 KW cell regeneration; membrane protein; signal transduction-related protein;  
 KW transcription-related protein; osteoporosis; neurological disease;  
 KW cancer; tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI308459-A2.  
 XX  
 PD 07-MAY-2003.  
 XX  
 XX 28-MAR-2002; 2002EP-00007401.  
 XX  
 XX 05-NOV-2001; 2001JP-00379298.  
 XX  
 XX 25-JAN-2002; 2002US-00350978.  
 XX  
 XX (HELI-) HELIX RES INST.  
 XX  
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
 XX  
 XX WPI; 2003-450961/43.  
 XX  
 XX N-PSDB; ADB62950.  
 XX  
 XX New polynucleotides and polypeptides, useful for developing a diagnostic  
 XX marker or medicines for regulation of their expression and activity, or  
 XX as targets of gene therapy.  
 XX  
 XX Claim 1; Page; 222pp; English.  
 XX  
 XX The invention discloses a polynucleotide comprising a sequence selected  
 XX from 1970 fully defined nucleotide sequences which encode novel  
 XX polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 XX or its partial peptide, an antibody binding to the polypeptide or peptide  
 XX of the polynucleotide, immunologically assaying the polypeptide or  
 XX peptide of the polynucleotide by contacting the polypeptide or peptide  
 XX with the antibody of the encoded protein, and observing the binding  
 XX between the two a transformant carrying the polynucleotide in an  
 XX expressible manner and an antisense polynucleotide. The oligonucleotide  
 XX is useful as a primer for synthesising the polynucleotide, or as a probe  
 XX for detecting the polynucleotide. The polynucleotides and encoded  
 XX proteins are useful as pharmaceutical agents and many disease-related  
 XX genes may be included in them, for developing a diagnostic marker or  
 XX medicines for regulation of their expression and activity, or as targets



CC for alleviating the symptoms associated with diseases involving chronic  
 CC immune cell dysfunction or to treat autoimmune diseases such as systemic  
 CC lupus erythematosus, rheumatoid arthritis, multiple sclerosis, diabetes,  
 CC immune thrombocytopenic purpura and psoriasis, chronic inflammatory  
 CC disease such as inflammatory bowel disease (Crohn's disease and  
 CC ulcerative colitis), Grave's disease, Hashimoto's thyroiditis and  
 CC diabetes mellitus. They are also useful as immunosuppressive agents for  
 CC bone marrow and organ transplantation or to prolong graft survival. B7-L  
 CC molecules are also useful for diagnosis and treatment of diseases  
 CC involving abnormal cell proliferation, including arteriosclerosis and  
 CC vascular restenosis. Antagonists of B7-L polypeptides are useful for  
 CC alleviation of toxic shock syndrome or allosensitisation due to blood  
 CC transfusions, and for treatment of allergy, asthma and hypersensitivity  
 CC reactions, nephropathies (e.g. glomerulonephritis), skin disorders  
 CC (pemphigus and pemphigoid), endocrinopathies (Grave's disease), various  
 CC pneumopathies (extrinsic alveolitis), vasculopathies, coeliac disease,  
 CC anaemia, thrombocytopenias, Guillain-Barre syndrome and myasthenia  
 CC gravis, and lymphoproliferative disorders such as multiple myeloma. The  
 CC present sequence represents the amino acid sequence of human B7-L<sub>h2</sub>  
 XX

SQ Sequence 386 AA;

#### Alignment Scores:

Pred. No.: 2,24e-239 Length: 386  
 Score: 257.00 Matches: 357  
 Percent Similarity: 99.72% Conservative: 0  
 Best Local Similarity: 99.72% Mismatches: 1  
 Query Match: 66.07% Indels: 0  
 DB: 5 Gaps: 0

US-09-729-264-1 (1-1175) x AAU75541 (1-386)

QY 99 GTCTGAGGGCTCCCAAGGCTCGCTTCAACTGCACCGCTCTCCAGGGCTGGAAGCTCATC 158  
 DB 29 ValLeuLysGlySerGlnAlaArgPheAsnCysThrValSerGlnGlyTrpLysLeuLe 48  
 QY 159 ATGTGGGCTCTCAGTGACATGCTGCTGAAGCTGACGGCCATGAGGCCATCATCACC 218  
 DB 49 MetTrpAlaLeuSerAspMetValValLeuSerValArgProMetGluProIleThr 68  
 QY 219 AATGACCGCTTCACTCTCAGAGTACGACCGGGGGAAGTTCACCTCGGAGATCATC 278  
 DB 69 AsnAspArgPheThrSerGlnArgTyrAspGlnGlyAsnPheThrSerGluMetIle 88  
 QY 279 ATCCACAATGTGGAGCCAGTGATTTCGGGGAACATCAGATCGACCTCCAGAACAGTCGC 338  
 DB 89 IleHisAsnValGluProSerAspSerGlyAsnIleArgCysSerLeuGlnAsnSerArg 108  
 QY 339 CTGCATCGATCTGCTTACCTTACCGTCCAAAGTTATGGAGAGCTGTTTCATCCAGTGT 398  
 DB 109 LeuHisGlySerAlaTyrLeuThrValGlnValMetGlyGluLeuPheIleProSerVal 128  
 QY 399 AATCTTCTAGTCGCTGAGATGAACCTGTGAAGTTACTTGTCTACCTCCACACTGCAC 458  
 DB 129 AsnLeuValValAlaGluAsnGluProCysGluValThrCysLeuProSerHisTrpThr 148  
 QY 459 CGGCTCCCGGATATTCTCGGAGCTCGGTCTCTCGTTCAGCCATTCAGCTATTATTTT 518  
 DB 149 TrpLeuProAspIleSerTrpGluLeuGlyLeuLeuValSerHisSerSerTyrTrpPhe 168  
 QY 519 GTTCCGGAGCCAGCCAGCTTCAAGTGCAGTGAGATCTCGGCTCGACCCACAGCAGC 578  
 DB 169 ValProGluProSerAspLeuGlnSerAlaValSerIleLeuAlaLeuThrProGlnSer 188  
 QY 579 AATGGGACTTGTACTTCGCTGGCTTACCTGGAGAGCTTGAAGCCCGCAAGTCTGCAACT 638  
 DB 189 AsnGlyThrLeuThrCysValAlaThrTrpLysSerLeuLysAlaArgLysSerAlaThr 208  
 QY 639 GTAATCTACTGTGATTTCGGTGTCCCAAGACACTGGAGGTGGTATTAATATTCAGGT 698  
 DB 209 ValAsnLeuThrValIleArgCysProGlnAspThrGlyGlyIleAsnIleProGly 228  
 QY 699 GTATTATCAAGTTTACCGAGTTTAGTGTTCATTCGCTACTTGGGCGCAAGTTGGACT 758

DB 229 ValLeuSerSerLeuProSerLeuGlyPheSerLeuProThrTrpGlyLysValGlyLeu 248  
 QY 759 GGACTAGCAGGACCATGCTTCTCAGCCGACGCTACTCTTCAATACGCTGCTGCTGC 818  
 DB 249 GlyLeuAlaGlyThrMetLeuLeuThrProThrCysThrLeuThrIleArgCysCys 268  
 QY 819 TGCCGCGCTGTTGTTGTGGCTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 878  
 DB 269 CysArgArgArgCysCysGlyCysAsnCysCysArgCysCysPheCysCysArgArg 288  
 QY 879 AAAAGAGATTTCGTATTTCATTTTCAAAAGAAATCTGAAAAGAGAGACAAACAAGAA 938  
 DB 289 LysArgglyPheArgIleGlnPheGlnLysSerGluLysGluLysThrAsnLysGlu 308  
 QY 939 ACTGAGACAGAAAGTGGAAATGAAATCTCCGGCTTACAAATTCAGATGAACAAAGACCA 998  
 DB 309 ThrGluThrGluSerGlyAsnGluAsnSerGlyTyrAsnSerAspGluGlnLysThr 328  
 QY 999 GACACCGCTTCTCTCCCTCCCAATCTGTGAATCCAGTGATCTGTAACAAAGAAACAGT 1058  
 DB 329 AspThrAlaSerLeuProLysSerCysGluSerSerAspProGluGlnArgAsnSer 348  
 QY 1059 AGCTGTGGCCCTCTCTCAGCAGCGGGCTGATCAACGTCACCCAGGCCAGCAAGTCATCCA 1118  
 DB 349 SerCysGlyProProHisGlnArgAlaAspGlnArgProProArgProAlaSerHisPro 368  
 QY 1119 CAGGCTTCTTTTAAATCTGGCCAGTCCCTGAGAAGTCAAGTAAATCAACTGTAGTA 1172  
 DB 369 GlnAlaSerPheAsnLeuAlaSerProGluLysValSerAsnThrThrValVal 386

#### RESULT 5

ABG28169  
 ID ABG28169 standard; protein; 463 AA.  
 XX  
 AC ABG28169;  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #28160.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 XX 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Drmanac RT, Liu C, Tang YT;  
 PI  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS923356.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 PT

PS Claim 20; SEQ ID NO 58528; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used



CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 463 AA;

## Alignment Scores:

Pred. No.:	7, 77e-226	Length:	463
Score:	243.00	Matches:	243
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	62.47%	Indels:	0
DB:	4	Gaps:	0

US-09-729-264-1 (1-1175) x ABG28169 (1-463)

QY	159	ATGTGGCTCTCAGTGACATGGTGTGCTAAGCGTCAGGCCCATGAGCCCATCATCACC	218
DB	1	MetTrpAlaLeuSerAspMetValValLeuSerValArgProMetGluProIleThr	20
QY	219	AATGACCGCTTCACCTCTCAGAGGTACGACAGCGCGGGAACCTCACCTCGAGATGATC	278
DB	21	AsnAspArgPheThrSerGlnArgTyrAspGlnGlyAsnPheThrSerGluMetIle	40
QY	279	ATCCACAATGTGGAGCCAGTCGATTCGGGGAACATCAGATGCGCTCCAGAACAGTCGC	338
DB	41	IleHisAsnValGluProSerAspSerGlyAsnIleArgCysSerIleuGlnAsnSerArg	60
QY	339	CTGATGATCTGCTTACCTTACCGTCCAGTATGAGTATGGAGAGCTGTTCAATCCAGTGT	398
DB	61	LeuHisGlySerAlaTyrLeuThrValGlnValMetGlyGluLeuPheIleProSerVal	80
QY	399	AATCTTGTAGTCGTGAGATCAACCTTGCTGAGTACTTGCTACTCCCTCACACTGGACC	458
DB	81	AsnLeuValAlaGluAsnGluProCysGluValThrCysLeuProSerHisTrpThr	100
QY	459	CGGCTCCCGGATATTCTCTGGAGCTCGCTCTCTCTGCTCAGCAATCAAGCTATTATTTT	518
DB	101	ArgLeuProAspIleSerTrpGluLeuGlyLeuLeuValSerHisSerTyrTrpPhe	120
QY	519	GTTCCGGAGCCAGGACCTTCAAGTCAGTGCATCTCTGCTCTGCTTCAACCCACAGACC	578
DB	121	ValProGluProSerAspSerGlnSerAlaValSerIleLeuAlaLeuThrProGlnSer	140
QY	579	AATGGGACTTTCACCTGCTGCTACCTGAGAGCTGAGGCGCGGCAAGCTCTCAACT	638
DB	141	AsnGlyThrLeuThrCysValAlaThrTrpIysSerLeuLeuAlaArgIysSerAlaThr	160
QY	639	GTAATCTCACTGTGATTCGGTGTGCTCCCAAGACACTGGAGGTGGTATTAAATTCACAGT	698
DB	161	ValAsnLeuThrValIleArgCysProGlnAspThrGlyGlyIleAsnIleProGly	180
QY	699	GTAATCAAGTTTACCGAGTTTAGGTTTTCATTCCTTACTTGGCGGCAAGTTCGACTT	758
DB	181	ValLeuSerSerLeuProSerLeuGlyPheSerLeuProThrTrpGlyIysValGlyLeu	200
QY	759	GGACTAGCAGGACCATCTCTGACGCGAGCTGTACTTACATAGCTGTGCTGTC	818
DB	201	GlyLeuAlaGlyThrMetLeuLeuThrProThrCysThrLeuThrIleArgCysCys	220

QY	819	TGCCCGCGTCTGTTGTTGGCTGCAACTGCTGCTGCGTGTGTTCTCTGCTGTAGAGA	878
DB	221	CysAspArgGlyGlyCysGlyCysArgCysArgCysPheCysArgArg	240
QY	879	AAAAGAGGA 887	
DB	241	LyAsArgGly 243	
RESULT 6			
AAU75543			
ID	AAU75543	standard; protein; 377 AA.	
XX			
AC	AAU75543;		
XX			
DT	23-APR-2002	(first entry)	
XX			
DE	Human B7-like protein, B7-L_h4.		
XX			
KW	Human; B7-like protein; B7-L; antiinfertility; gynaecological;		
KW	antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;		
KW	antiinflammatory; dermatological; antipsoriatic; neuroprotective;		
KW	antidiabetic; haemostatic; antithyroid; antileuk; antiallergic;		
KW	antialsthatic; nephrotropic; antibacterial; virucide; tumour; cancer;		
KW	reproductive disorder; graft versus host disease; autoimmune disease;		
KW	toxic shock syndrome; allergy; nephropathy; skin disorder;		
KW	endocrinopathy; lymphoproliferative disorder.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200200710-A2.		
XX			
PD	03-JAN-2002.		
XX			
PF	28-JUN-2001; 2001WO-US020719.		
XX			
PR	28-JUN-2000; 2000US-0214512P.		
PR	28-NOV-2000; 2000US-00729264.		
XX	(AMGE-) AMGEN INC.		
PA			
PI	Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;		
XX			
DR	WPI; 2002-130881/17.		
DR	N-PSDB; ASK13031.		
XX			
PT	New B7-like polypeptides, polynucleotides and their modulators, useful		
PT	for diagnosing, preventing and treating reproductive, immune and		
PT	proliferative disorders, e.g. cancer and arteriosclerosis.		
XX			
PS	Claim 13; Fig 4; 135pp; English.		
CC			
CC	The invention relates to an isolated B7-like (B7-L) polypeptide (I). The		
CC	polypeptide, polynucleotide encoding it and antibody against (I) are		
CC	useful for treating B7-like polypeptide-related disease, disorders or		
CC	conditions including reproductive disorders (e.g. infertility,		
CC	miscarriage, preterm labour and delivery and endometriosis) and		
CC	proliferative disorders. Antibodies, soluble proteins comprising		
CC	extracellular domains and other regulators of B7-L polypeptides are		
CC	useful for enhancing the immune response to tumours. (I) plays a role in		
CC	growth and maintenance of cancer cells based on the observation of		
CC	seminal vesicle hyperplasia in transgenic mice overexpressing B7-L		
CC	polypeptide. Hence modulators of (I) are useful for the treatment of		
CC	cancer including seminal vesicle cancer, lung, brain, breast, ovarian,		
CC	testicular cancer and cancers of haematopoietic system. B7-L polypeptide		
CC	pathway can be manipulated to regulate cytotoxic T-lymphocyte response in		
CC	allograft transplantation, graft versus host disease, T-cell dependent B-		
CC	cell mediated diseases and autoimmune diseases. B7-L molecules are useful		
CC	for alleviating the symptoms associated with diseases involving chronic		
CC	immune cell dysfunction or to treat autoimmune diseases such as systemic		
CC	lupus erythematosus, rheumatoid arthritis, multiple sclerosis, diabetes,		
CC	immune thrombocytopenic purpura and psoriasis, chronic inflammatory		
CC	disease such as inflammatory bowel disease (Crohn's disease and		

CC ulcerative colitis), Grave's disease, Hashimoto's thyroiditis and  
CC diabetes mellitus. They are also useful as immunosuppressive agents for  
CC bone marrow and organ transplantation or to prolong graft survival. B7-L  
CC molecules are also useful for diagnosis and treatment of diseases  
CC involving abnormal cell proliferation, including arteriosclerosis and  
CC vascular restenosis. Antagonists of B7-L polypeptides are useful for  
CC alleviation of toxic shock syndrome or allersensitisation due to blood  
CC transfusions, and for treatment of allergy, asthma and hypersensitivity  
CC reactions, nephropathies (e.g. glomerulonephritis), skin disorders  
CC (pemphigus and pemphigoid), endorinopathies (Grave's disease), various  
CC pneumopathies (extrinsic alveolitis), vasculopathies, coeliac disease,  
CC anaemia, thrombocytopenias, Guillain-Barre syndrome and myasthenia  
CC gravis, and lymphoproliferative disorders such as multiple myeloma. The  
CC present sequence represents the amino acid sequence of human B7-L<sub>h4</sub>  
XX  
XX  
SQ Sequence 377 AA;

## Alignment Scores:

Pred. No.: 5.84e-192 Length: 377  
Score: 208.00 Matches: 208  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 53.47% Indels: 0  
DB: Gaps: 0

US-09-729-264-1 (1-1175) x AAU75543 (1-377)

QY 264 ACCTCGGAGATGATCATCCAAATGTGGAGCCAGTATTGGGGAACATCAGATGCAGC 323  
Db 84 ThrSerGluMetIleIleHisAsnValGluProSerAspSerGlyAsnIleArgCysSer 103  
QY 324 CTCGAGAACAGTCGCTCATGTGATCTGCTTACCTTACCGTCCCAAGTTATGGGAGAGCTG 383  
Db 104 LeuGlnAsnSerArgLeuHisGlySerAlaIleLeuThrValGlnValMetGlyGluLeu 123  
QY 384 TTGATTCCTCCAGTGTAACTTGTAGTCGTGAGAAATGAACCTTGTGAGTACTTGTCTA 443  
Db 124 PheIleProSerValAsnLeuValAlaGluAsnGluProCysGluValThrCysLeu 143  
QY 444 CCCTCACATGGAGCCGGCTCCGGATATTTCCGTGGAGCTCGCTTCCGTGTCAGGCAT 503  
Db 144 ProSerHisTrpThrArgLeuProAspIleSerTrpGluLeuGlyLeuValSerHis 163  
QY 504 TCAAGCTATTATTTTGTTCGGAGCCAGCAGCACTTCAAAGTGCAGTGCAGTCTCGCT 563  
Db 164 SerSerTyrrPheValProGluProSerAspLeuGlnSerAlaValSerIleLeuAla 183  
QY 564 CTGACCCGACAGCAATGGAGCTTTGACTTGCCTGGCTACCTGGAAGAGCCTGAAGCC 623  
Db 184 LeuThrProGlnSerAsnGlyThrLeuThrCysValAlaIleThrTrpLysSerLeuLysAla 203  
QY 624 CGCAAGTCTGCACTGTAAATCTCAGTGTGATTCGGTGTCCCAAGACACTGGAGGTGGT 683  
Db 204 ArgLysSerAlaThrValAsnLeuThrValIleArgCysProGlnAspThrGlyGlyGly 223  
QY 684 ATTAATATTCAGGTGATTATCAAGTTTACCGAGTTTAGGTTTTCATTCCTGCTACTGG 743  
Db 224 IleAsnIleProGlyValLeuSerSerLeuProSerLeuGlyPheSerLeuProThrTrp 243  
QY 744 GGCAGAGTCTGACTTGGACTAGCAGCAGCAGCAATGCTTCTGACGCGAGCTGTACTTACA 803  
Db 244 GlyLysValGlyLeuGlyLeuAlaGlyThrMetLeuLeuThrProThrCysThrLeuThr 263  
QY 804 ATACGCTGCTGCTGCTCCGCGCTGTGTTGTTGTGGCTGCAACTGCTGCTGCGCTGTGTTGT 863  
Db 264 IleArgCysCysCysCysArgArgArgCysCysGlyCysAsnCysCysCysArgCysCys 283  
QY 864 TTCTGCTGTAGAGAAAGAGGA 887  
Db 284 PheCysCysArgArgLysArgGly 291  
RESULT 7  
AAM20679

ID AAM20679 standard; protein; 78 AA.  
XX  
AC AAM20679;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Peptide #7113 encoded by probe for measuring cervical gene expression.  
XX  
KW Probe; human; microarray; gene expression; cervical epithelial cell;  
KW cervical cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157278-A2.  
XX  
XX 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000670.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-498901/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human cervical epithelial cells.  
XX  
PS Claim 27; SEQ ID NO 25505; 487pp; English.  
XX  
XX The present invention relates to human single exon nucleic acid probes  
XX (SENPs: see AAI10069-AA128459). The present sequence is a peptide encoded  
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
XX can be used to produce a single exon microarray, which can be used for  
XX measuring human gene expression in a sample derived from human cervical  
XX epithelial cells. By measuring gene expression, the probes are therefore  
XX useful in grading and/or staging of diseases of the cervix, notably  
XX cervical cancer. Note: The sequence data for this patent did not form  
XX part of the printed specification, but was obtained in electronic format  
XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 78 AA;  
Alignment Scores:  
Pred. No.: 0.0309 Length: 78  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.16% Indels: 0  
DB: Gaps: 0  
US-09-729-264-1 (1-1175) x AAM20679 (1-78)  
QY 842 GCAGCCACACAGCAGCGCGGCGGAGCAGCAGCAGCG 807  
Db 4 AlaAlaThrThrThrAlaAlaAlaAlaAla 15  
RESULT 8  
ABB42042  
ID ABB42042 standard; peptide; 78 AA.  
XX  
AC ABB42042;  
XX  
XX 04-FEB-2002 (first entry)  
XX

DE Peptide #9548 encoded by human foetal liver single exon probe.  
XX Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX Homo sapiens.  
XX WO200157277-A2.  
XX 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US000669.  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human fetal liver.  
XX Claim 27; SEQ ID NO 34677; 639pp + Sequence Listing; English.  
XX The invention relates to a single exon nucleic acid probe for measuring  
XX human gene expression in a sample derived from human foetal liver. The  
XX single exon nucleic acid probes may be used for predicting, measuring and  
XX displaying gene expression in samples derived from human fetal liver. The  
XX present sequence is a peptide encoded by a single exon nucleic acid probe  
XX of the invention. Note: The sequence data for this patent did not form  
XX part of the printed specification, but was obtained in electronic format  
XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 78 AA;  
SQ  
Alignment Scores:  
Pred. No.: 0.0309 Length: 78  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.18% Indels: 0  
DB: 4 Gaps: 0  
US-09-729-264-1 (1-1175) x ABB42042 (1-78)  
QY 842 GCAGCCACACAGCAGCGGCGCAGCAGCG 807  
Db 4 AlaAlaThrThrThrAlaAlaAlaAlaAla 15  
RESULT 9  
AAM35844  
ID AAM35844 standard; protein; 78 AA.  
AC AAM35844;  
XX 17-OCT-2001 (first entry)  
XX Peptide #9881 encoded by probe for measuring placental gene expression.  
XX Probe; microarray; human; placenta; antenatal diagnosis;  
XX genetic disorder.  
XX Homo sapiens.  
XX WO200157272-A2.  
XX

PD 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US000663.  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488897/53.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human placenta.  
XX Claim 27; SEQ ID NO 36113; 654pp; English.  
XX The present invention relates to single exon nucleic acid probes (SENP;  
XX see AAI31315-AAI57546). The present sequence is a peptide encoded by one  
XX such probe. The probes are useful for producing a microarray for  
XX predicting, measuring and displaying gene expression in samples derived  
XX from human placenta. The probes are useful for antenatal diagnosis of  
XX human genetic disorders  
XX Sequence 78 AA;  
SQ  
Alignment Scores:  
Pred. No.: 0.0309 Length: 78  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.18% Indels: 0  
DB: 4 Gaps: 0  
US-09-729-264-1 (1-1175) x AAM35844 (1-78)  
QY 842 GCAGCCACACAGCAGCGGCGCAGCAGCG 807  
Db 4 AlaAlaThrThrThrAlaAlaAlaAlaAla 15  
RESULT 10  
ABB25654  
ID ABB25654 standard; protein; 78 AA.  
XX ABB25654;  
XX 23-JAN-2002 (first entry)  
XX Protein #7653 encoded by probe for measuring heart cell gene expression.  
XX Human; gene expression; heart; microarray; vascular system;  
XX cardiovascular disease; hypertension; cardiac arrhythmia;  
XX congenital heart disease.  
XX Homo sapiens.  
XX WO200157274-A2.  
XX 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US000666.  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX

PR 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488899/53.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts.  
XX  
XX Claim 15; SEQ ID NO 27424; 530pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes for  
XX measuring human gene expression in a sample derived from human heart (see  
XX ABA21535-ABM41105). The present sequence is a protein encoded by one such  
XX probe. The probes may be used for predicting, measuring and displaying  
XX gene expression in samples derived from the human heart via microarrays.  
XX By measuring gene expression, the probes are useful for predicting,  
XX diagnosing, grading, staging, monitoring and prognosing diseases of the  
XX human heart and vascular system e.g. cardiovascular disease,  
XX hypertension, cardiac arrhythmias and congenital heart disease. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 78 AA;

Alignment Scores:  
Pred. No.: 0.0309 Length: 78  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.16% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x ABB25654 (1-78)

QY 842 GCAGCCACACACGCGCGGCGAGCAGCAGCG 807  
Db 4 AlaAlaThrThrThrAlaAlaAlaAlaAla 15

RESULT 11  
AAM75735  
ID AAM75735 standard; protein; 78 AA.

XX AAM75735;  
AC  
XX  
XX 06-NOV-2001 (first entry)  
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 36041.  
XX  
XX Human; bone marrow expressed exon; gene expression analysis; probe;  
XX microarray; cancer; leukaemia; lymphoma; myeloma.  
XX Homo sapiens.  
XX  
XX WO200157276-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000668.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human bone marrow.  
XX  
XX Example 4; SEQ ID NO 36041; 658pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukaemia and myeloma. The present sequence is a  
XX protein encoded by one of the probes of the invention  
XX  
XX Sequence 78 AA;  
XX  
XX Alignment Scores:  
Pred. No.: 0.0309 Length: 78  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.16% Indels: 0  
DB: 4 Gaps: 0  
XX  
XX US-09-729-264-1 (1-1175) x AAM75735 (1-78)  
XX  
XX QY 842 GCAGCCACACACGCGCGGCGAGCAGCAGCG 807  
XX Db 4 AlaAlaThrThrThrAlaAlaAlaAlaAla 15  
XX  
XX RESULT 12  
XX AAM62923  
XX ID AAM62923 standard; protein; 78 AA.  
XX  
XX AAM62923;  
AC  
XX  
XX 05-NOV-2001 (first entry)  
XX  
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 35028.  
XX  
XX Human; brain expressed exon; gene expression analysis; probe; microarray;  
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
XX Homo sapiens.  
XX  
XX WO200157275-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000667.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains.  
XX  
XX Example 4; SEQ ID NO 35028; 650pp + Sequence Listing; English.  
XX

CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention

XX SQ Sequence 78 AA;

Alignment Scores:  
Pred. No.: 0.0309 Length: 78  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.16% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x AAM62923 (1-78)

QY 842 GCAGCCACACACGCGCGCAGCAGCAGCG 807  
D 4 AlaAlaThrThrThrAlaAlaAlaAlaAla 15

RESULT 13

ABG57473  
ID ABG57473 standard; peptide; 78 AA.

AC ABG57473;

DT 25-FEB-2003 (first entry)

DE Human liver peptide, SEQ ID No 36121.

XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
KW hypercholesterolaemia; coronary heart disease.

XX Homo sapiens.

XX WO200157273-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000664.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488898/53.

XX Claim 27; SEQ ID NO 36121; 658pp; English.

CC The invention relates to a single exon nucleic acid probe (SEN) (I) for  
CC measuring human gene expression in a sample derived from human adult  
CC liver, comprising one of 13109 defined nucleotide sequences given in the  
CC specification (or complements/ fragments). The probe hybridises at high  
CC stringency to a nucleic acid molecule expressed in the human adult liver.  
CC (I) may be used for predicting, measuring and displaying gene expression  
CC in samples derived from human adult liver. The genes identified may be  
CC involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is

CC associated with coronary heart disease. ABG47348-ABG59930 represent human  
CC liver single exon encoded peptides of the invention. Note: The sequence  
CC information for this patent does not appear in the printed specification  
CC but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 78 AA;

Alignment Scores:  
Pred. No.: 0.0309 Length: 78  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 3.16% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x ABG57473 (1-78)

QY 842 GCAGCCACACACGCGCGCAGCAGCAGCG 807  
D 4 AlaAlaThrThrThrAlaAlaAlaAlaAla 15

RESULT 14

ABG45219  
ID ABG45219 standard; peptide; 78 AA.

AC ABG45219;

DT 19-AUG-2002 (first entry)

DE Human peptide encoded by genome-derived single exon probe SEQ ID 34884.

XX Human; single exon probe; asthma; lung cancer; COPD; ILD;

XX chronic obstructive pulmonary disease; interstitial lung disease;

XX familial idiopathic pulmonary fibrosis; neurofibromatosis;

XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;

XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

XX primary ciliary dyskinesia; pulmonary hypertension;

XX hyaline membrane disease.

XX Homo sapiens.

XX WO200186003-A2.

XX 15-NOV-2001.

XX 30-JAN-2001; 2001WO-US000665.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2002-114183/15.

XX Claim 27; SEQ ID NO 34884; 634pp; English.

CC Spatially-addressable set of single exon nucleic acid probes, used to  
CC measure gene expression in human lung samples.

XX Claim 27; SEQ ID NO 34884; 634pp; English.

CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their

complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 78 AA;

Alignment Scores:  
 Pred. No.: 0.0309 Length: 78  
 Score: 12.00 Matches: 12  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 3.16% Indels: 0  
 DB: 5 Gaps: 0

US-09-729-264-1 (1-1175) x ABG45219 (1-78)

QY 842 GCAGCCACACAGCAGCGCGCAGCAGCAGCG 807  
 Db 4 AlaAlaThrThrThrAlaAlaAlaAlaAla 15

RESULT 15  
 AAU07370

ID AAU07370 standard; protein; 524 AA.

XX

AC AAU07370;

XX 18-DEC-2001 (first entry)

XX G protein-coupled receptor.

XX Human; mental disorder; thyroid disease; renal failure; anorexia;  
 KW inflammatory condition; Crohn's disease; rheumatoid arthritis; HIV;  
 KW autoimmune disorder; schizophrenia; migraine; stroke; dementia; obesity;  
 KW depression; Parkinson's disease; Alzheimer's disease; viral infection;  
 KW Huntington's disease; human immunodeficiency virus; type 2 diabetes;  
 KW anorexia; hypotension; hypertension; thrombosis; myocardial infarction;  
 KW atherosclerosis; cancer; sexual dysfunction; G protein-coupled receptor;  
 KW nGPCR.

XX Homo sapiens.

OS

XX W02001.62924-A2.

PN

XX 30-AUG-2001.  
 PD  
 XX  
 XX 23-FEB-2001; 2001WO-US005989.  
 PF  
 XX  
 XX 24-FEB-2000; 2000US-0184602P.  
 PR  
 XX 24-FEB-2000; 2000US-0184604P.  
 PR  
 XX 24-FEB-2000; 2000US-0184606P.  
 PR  
 XX 24-FEB-2000; 2000US-0184689P.  
 PR  
 XX 24-FEB-2000; 2000US-0184690P.  
 PR  
 XX 24-FEB-2000; 2000US-0184710P.  
 PR  
 XX 24-FEB-2000; 2000US-0184712P.  
 PR  
 XX 24-FEB-2000; 2000US-0184715P.  
 PR  
 XX 24-FEB-2000; 2000US-0184716P.  
 PR  
 XX 24-FEB-2000; 2000US-0184725P.  
 PR  
 XX 24-FEB-2000; 2000US-0184822P.  
 PR  
 XX (PHAA ) PHARMACIA & UPJOHN CO.  
 PA  
 XX  
 XX Vogel G, Wood LS, Parodi LA, Lind P;  
 PI  
 XX WPI; 2001-570632/64.  
 DR

XX Novel nucleic acid and encoded nGPCR-x, used to screen for compounds for use in the treatment of mental disorders, such as Alzheimer's disease, or Parkinson's disease.

XX Claim 31; Page 178-179; 263pp; English.

XX The invention relates to novel isolated human G protein-coupled receptors (nGPCR-x). The nGPCR-x can be used for screening compounds which can be used to treat mental disorders, thyroid disease, renal failure, inflammatory conditions such as Crohn's disease, rheumatoid arthritis, autoimmune disorders, schizophrenia, migraine, stroke, dementia, depression, Parkinson's disease, Alzheimer's disease, and Huntington's disease. They may also be used for treating viral infections such as human immunodeficiency virus (HIV), type 2 diabetes, obesity, anorexia, hypotension, hypertension, thrombosis, myocardial infarction, atherosclerosis, cancer, and sexual dysfunction. AAU25617-AAU25726 represent the amino acid sequences of novel human G protein-coupled receptors, nGPCR-2031 to nGPCR-2140 respectively, as described in the invention

XX Sequence 524 AA;

Alignment Scores:  
 Pred. No.: 0.24 Length: 524  
 Score: 11.00 Matches: 11  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.89% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x AAU07370 (1-524)

QY 839 GCACACACAGCAGCGCGCAGCAGCAGCG 807

Db 432 AlaThrThrThrAlaAlaAlaAlaAla 442

RESULT 16

AB865262

ID AB865262 standard; protein; 1805 AA.

XX

AC AB865262;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 22578.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

OS

XX PN WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX PF 23-MAR-2001; 2001WO-US009231.  
 XX PR 23-MAR-2000; 2000US-0191637P.  
 XX PR 11-JUL-2000; 2000US-00614150.  
 XX PA (PEKE ) PE CORP NY.  
 XX PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX DR WPI; 2001-656860/75.  
 XX DR N-PSDB; ABL09365.  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX PS Disclosure; SEQ ID NO 22578; 21pp + Sequence Listing; English.  
 XX CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB163051), expressed DNA  
 CC sequences (AB161840-AB16175) and the encoded proteins (AB161773-  
 CC AB161702). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 1805 AA;

Alignment Scores:  
 Pred. No.: 0.213 Length: 1805  
 Score: 11.00 Matches: 11  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.83% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x ABB65262 (1-1805)

QY 808 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
 Db 27 AlaAlaAlaAlaAlaAlaValValValAla 37

RESULT 17

AAV65872  
 ID AAV65872 standard; peptide; 85 AA.

XX AC AAV65872;

XX DT 10-FEB-2000 (first entry)

XX DE n-myc mutant peptide 1.

XX KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;  
 KW mutein.

XX OS Homo sapiens.  
 XX OS Synthetic.

XX PN WO9958552-A2.

XX PD 18-NOV-1999.

XX PF 03-MAY-1999; 99WO-NO000143.

XX PR 08-MAY-1998; 98NO-00002097.

XX PA (NHVD ) NORSK HYDRO AS.  
 XX PI Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;  
 XX DR WPI; 2000-039064/03.  
 XX PT New peptides derived from genes with frameshift mutations, used to  
 PT develop products for the treatment and prophylaxis of cancers.  
 XX PS Claim 13; Page 29; 166pp; English.

XX CC Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a  
 CC frameshift mutation in a gene from a cancer cell. The peptides are  
 CC characterised in that they: (i) are at least 8 amino acids long and a  
 CC fragment of a mutant protein arising from a frameshift mutation in a gene  
 CC of a cancer cell; (ii) consist of at least one amino acid of the mutant  
 CC part of a protein sequence encoded by the gene; (iii) comprise 0-10 amino  
 CC acid from the carboxyl terminus of the normal part of the protein  
 CC sequence preceding the amino terminus of the mutant sequence and may  
 CC further extend to the carboxyl terminus of the mutant part of the protein  
 CC as determined by a new stop codon generated by the frameshift mutation;  
 CC and (iv) induce, either in their full lengths or after processing by an  
 CC antigen presenting cell (APC), T cell responses. The genes that the  
 CC peptides are derived from, are characterised as susceptible to frameshift  
 CC mutation by having a mono nucleoside base repeat sequence of at least 5  
 CC residues, or a di-nucleoside base repeat sequence of at least 4 di-  
 CC nucleoside base units. The peptides are created by the addition or  
 CC deletion of 1 or 2 nucleoside base residues from the repeat sequence. The  
 CC novel peptides can elicit T cell responses and toxicity against tumours  
 CC and cancer cells carrying genes with frameshift mutations. The novel  
 CC peptides and DNA sequences can be used for the preparation of a  
 CC composition for the treatment or prophylaxis of cancer

XX SQ Sequence 85 AA;

Alignment Scores:  
 Pred. No.: 2.64 Length: 85  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.57% Indels: 0  
 DB: 3 Gaps: 0

US-09-729-264-1 (1-1175) x AAY65872 (1-85)

QY 124 TCRACTGCACGCTCTCCAGCGCTGGAGC 153  
 Db 25 SerThrAlaProSerProArgAlaGlySer 34

RESULT 18

AAV65874

ID AAV65874 standard; peptide; 94 AA.

XX AC AAV65874;

XX DT 10-FEB-2000 (first entry)

XX DE n-myc mutant peptide 3.

XX KW Human; frameshift mutant; T cell response; tumour; treatment; cancer;  
 KW mutein.

XX OS Homo sapiens.  
 XX OS Synthetic.

XX PN WO9958552-A2.

XX PD 18-NOV-1999.

XX PF 03-MAY-1999; 99WO-NO000143.

XX PR 08-MAY-1998; 98NO-00002097.

XX (NH2D ) NORSEK HYDRO AS.  
XX Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;  
XX WPI; 2000-039064/03.  
XX  
XX New peptides derived from genes with frameshift mutations, used to  
XX develop products for the treatment and prophylaxis of cancers.  
XX  
XX Claim 13; Page 29; 166pp; English.  
XX  
XX Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a  
XX frameshift mutation in a gene from a cancer cell. The peptides are  
XX characterised in that they: (i) are at least 8 amino acids long and a  
XX fragment of a mutant protein arising from a frameshift mutation in a gene  
XX of a cancer cell; (ii) consist of at least one amino acid of the mutant  
XX part of a protein sequence encoded by the gene; (iii) comprise 0-10 amino  
XX acid from the carboxyl terminus of the normal part of the protein  
XX sequence preceding the amino terminus of the mutant sequence and may  
XX further extend to the carboxyl terminus of the mutant part of the protein  
XX as determined by a new stop codon generated by the frameshift mutation;  
XX and (iv) induce, either in their full lengths or after processing by an  
XX antigen presenting cell (APC), T cell responses. The genes that the  
XX peptides are derived from, are characterised as susceptible to frameshift  
XX mutation by having a mono nucleoside base repeat sequence of at least 5  
XX residues, or a di-nucleoside base repeat sequence of at least 4 di-  
XX nucleoside base units. The peptides are created by the addition or  
XX deletion of 1 or 2 nucleoside base residues from the repeat sequence. The  
XX novel peptides can elicit T cell responses and toxicity against tumours  
XX and cancer cells carrying genes with frameshift mutations. The novel  
XX peptides and DNA sequences can be used for the preparation of a  
XX composition for the treatment or prophylaxis of cancer  
XX  
SQ Sequence 94 AA;

Alignment Scores:  
Pred. No.: 2.62 Length: 94  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.57% Indels: 0  
DB: 3 Gaps: 0

US-09-729-264-1 (1-1175) x AAY65874 (1-94)

QY 124 TCACCTGCACCGCTCCAGGCTGGAAGC 153  
Db 34 SerThrAlaProSerProArgAlaGlySer 43

RESULT 19  
AAY65875  
ID AAY65875 standard; peptide; 95 AA.  
XX  
XX AAY65875;  
AC  
DT 10-FEB-2000 (first entry)  
XX  
XX n-myc mutant peptide 4.  
DE  
XX Human; frameshift mutant; T cell response; tumour; treatment; cancer;  
XX mutein.  
KW  
XX Homo sapiens.  
XX OS  
XX Synthetic.  
XX  
XX WO9958552-A2.  
FN  
XX 18-NOV-1999.  
XX  
XX 03-MAY-1999; 99WO-NO000143.  
XX  
XX 08-MAY-1998; 98NO-00002097.  
PR

XX (NH2D ) NORSEK HYDRO AS.  
XX Gaudernack G, Eriksen JA, Moller M, Gjertsen MK, Saeterdal I;  
XX WPI; 2000-039064/03.  
XX  
XX New peptides derived from genes with frameshift mutations, used to  
XX develop products for the treatment and prophylaxis of cancers.  
XX  
XX Claim 13; Page 29; 166pp; English.  
XX  
XX Peptides AAY65684-Y66142 are fragments of mutant proteins arising from a  
XX frameshift mutation in a gene from a cancer cell. The peptides are  
XX characterised in that they: (i) are at least 8 amino acids long and a  
XX fragment of a mutant protein arising from a frameshift mutation in a gene  
XX of a cancer cell; (ii) consist of at least one amino acid of the mutant  
XX part of a protein sequence encoded by the gene; (iii) comprise 0-10 amino  
XX acid from the carboxyl terminus of the normal part of the protein  
XX sequence preceding the amino terminus of the mutant part of the protein  
XX as determined by a new stop codon generated by the frameshift mutation;  
XX and (iv) induce, either in their full lengths or after processing by an  
XX antigen presenting cell (APC), T cell responses. The genes that the  
XX peptides are derived from, are characterised as susceptible to frameshift  
XX mutation by having a mono nucleoside base repeat sequence of at least 5  
XX residues, or a di-nucleoside base repeat sequence of at least 4 di-  
XX nucleoside base units. The peptides are created by the addition or  
XX deletion of 1 or 2 nucleoside base residues from the repeat sequence. The  
XX novel peptides can elicit T cell responses and toxicity against tumours  
XX and cancer cells carrying genes with frameshift mutations. The novel  
XX peptides and DNA sequences can be used for the preparation of a  
XX composition for the treatment or prophylaxis of cancer  
XX  
SQ Sequence 95 AA;

Alignment Scores:  
Pred. No.: 2.61 Length: 95  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.57% Indels: 0  
DB: 3 Gaps: 0

US-09-729-264-1 (1-1175) x AAY65875 (1-95)

QY 124 TCACCTGCACCGCTCCAGGCTGGAAGC 153  
Db 35 SerThrAlaProSerProArgAlaGlySer 44

RESULT 20  
ABB64563  
ID ABB64563 standard; protein; 2168 AA.  
XX  
XX ABB64563;  
AC  
DT 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 20481.  
DE  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
KW  
XX Drosophila melanogaster.  
XX OS  
XX WO200171042-A2.  
FN  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
XX  
XX 11-JUL-2000; 2000US-00614150.  
PR



XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX PS WPI; 2001-656860/75.  
XX DR N-PSDB; ABL08666.  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell  
XX PT interactions.  
XX PS Disclosure; SEQ ID NO 20481; 21pp + Sequence Listing; English.  
XX CC The invention relates to an isolated nucleic acid detection reagent  
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is  
XX CC useful in developmental biology and in elucidating cell signaling and  
XX CC cell-cell interactions in higher eukaryotes for the development of  
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention  
XX CC discloses genomic DNA sequences (ABLI16176-ABL30511), expressed DNA  
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
XX CC ABB72072). The sequence data for this patent did not form part of the  
XX CC printed specification, but was obtained in electronic format directly  
XX CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 2168 AA;  
Alignment Scores:  
Pred. No.: 1.94 Length: 2168  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.63% Indels: 0  
DB: 4 Gaps: 0  
US-09-729-264-1 (1-1175) x ABB64563 (1-2168)  
QY 836 ACAACGACGCGGCGAGCAGCAGCG 807  
DB 973 ThrThrThrAlaAlaAlaAlaAlaAla 982  
RESULT 21  
ABB69039  
ID ABB69039 standard; protein; 96 AA.  
AC ABB69039;  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster polypeptide SEQ ID NO 33909.  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX DR N-PSDB; ABL13142.  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX Disclosure; SEQ ID NO 33909; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABLI16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
XX ABB72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 96 AA;  
Alignment Scores:  
Pred. No.: 24.2 Length: 96  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.37% Indels: 0  
DB: 4 Gaps: 0  
US-09-729-264-1 (1-1175) x ABB69039 (1-96)  
QY 833 ACAACGACGCGGCGAGCAGCAGCG 807  
DB 20 ThrThrThrAlaAlaAlaAlaAlaAla 28  
RESULT 22  
ABB63572  
ID ABB63572 standard; protein; 191 AA.  
XX AC ABB63572;  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster polypeptide SEQ ID NO 17508.  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX DR N-PSDB; ABL07675.  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell  
XX PT interactions.  
XX Disclosure; SEQ ID NO 17508; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 191 AA;

Alignment Scores:  
Pred. No.: 22.7 Length: 191  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.37% Indels: 0  
DB: Gaps: 0

US-09-729-264-1 (1-1175) x ABB63572 (1-191)

QY 863 ACAACAGCGCAGCAGCGTTCAGCC 837  
ID ABB70426 standard; protein; 191 AA.  
XX  
AC ABB70426;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 38070.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
XX  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL14529.  
XX  
XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 38070; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 191 AA;

Alignment Scores:  
Pred. No.: 22.7 Length: 191  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.37% Indels: 0  
DB: Gaps: 0

US-09-729-264-1 (1-1175) x ABB70426 (1-191)

QY 842 GCAGCCACACACGACGGCGCAGCA 816  
ID ABB67144 standard; protein; 191 AA.  
XX  
AC ABB67144;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 28224.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
XX  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL11247.  
XX  
XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 28224; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 191 AA;

Alignment Scores:  
Pred. No.: 22.7 Length: 191  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.37% Indels: 0  
DB: Gaps: 0

US-09-729-264-1 (1-1175) x ABB67144 (1-191)

OY 863 ACAACAACGGCAGCAGCTTGCAGCC 837  
|||||  
Db 96 ThrThrAlaAlaValAlaAla 104

RESULT 25  
ADA48420  
ID ADA48420 standard; protein; 253 AA.  
XX  
AC ADA48420;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Rice protein conferring disease resistance in plants.  
XX  
KW disease resistance; pathogen tolerance; plant pathogen; plant; rice.  
XX  
OS Oryza sativa.  
XX  
PN WO2003000906-A2.  
XX  
PD 03-JAN-2003.  
XX  
PF 21-JUN-2002; 2002WO-IB002453.  
XX  
PR 22-JUN-2001; 2001US-0300112P.  
XX  
PR 26-SEP-2001; 2001US-0352277P.  
XX  
PR 22-MAR-2002; 2002US-0366535P.  
XX  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
PI Glazebrook J, Briggs S, Cooper B, Goff SA, Moughamer T;  
XX Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;  
XX  
DR WPI; 2003-184052/18.  
XX  
DR N-PSDB; ADA48419.  
XX  
PT New polynucleotide comprising a plant nucleotide sequence having an open  
XX reading frame that encodes a polypeptide associated with disease  
XX resistance, useful for conferring resistance or tolerance to a plant  
XX pathogen.  
XX  
PS Claim 10; SEQ ID NO 490; 299pp; English.  
XX  
CC The invention relates to a novel isolated polynucleotide comprising a  
XX plant nucleotide sequence having an open reading frame that encodes a  
XX polypeptide associated with disease resistance or its fragment having  
XX substantially the same activity as the full-length polypeptide. The  
XX polynucleotide of the invention is useful for conferring resistance or  
XX tolerance to a plant pathogen. The present sequence represents a protein  
XX conferring disease resistance used in the invention.  
XX  
SQ Sequence 253 AA;  
XX  
Alignment Scores:  
Pred. No.: 22.1 Length: 253  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
Db: 6 Gaps: 0

US-09-729-264-1 (1-1175) x ADA48420 (1-253)  
OY 820 GCCGCGCTGTGTGCTGCGTCACT 846  
|||||  
Db 36 AlaAlaValValValAlaAlaThr 44

RESULT 26  
ABU17237  
ID ABU17237 standard; protein; 258 AA.  
XX  
AC ABU17237;  
XX

XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #2764.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Acinetobacter baumannii.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
XX  
PR 06-SEP-2001; 2001US-00948993.  
XX  
PR 25-OCT-2001; 2001US-0342923P.  
XX  
PR 08-FEB-2002; 2002US-00072851.  
XX  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
XX  
DR N-PSDB; ACA21107.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 45161; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational  
XX drug discovery programs, or for screening homologous nucleic acids  
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
XX the target prokaryotic essential genes. Note: The sequence data for this  
XX patent did not form part of the printed specification, but was obtained  
XX in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 258 AA;  
XX  
Alignment Scores:  
Pred. No.: 22.1 Length: 258  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: Gaps: 0

US-09-729-264-1 (1-1175) x ABU17237 (1-258)

QY 1057 GTAGCTGTGGCCCTCTCACCAGCGG 1083

DB 6 ValAlaValAlaLeuLeuThrSerGly 14

RESULT 27

ADA36602

ID ADA36602 standard; protein; 269 AA.

XX AC ADA36602;

XX AC ADA36602;

DT 20-NOV-2003 (first entry)

XX AC

XX Acinetobacter baumannii protein #3763.

XX Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;  
KW plant biocontrol agent.

XX AC

XX Acinetobacter baumannii.

XX US6562958-B1.

PN 13-MAY-2003.

XX AC

XX 04-JUN-1999; 99US-00328352.

XX 09-JUN-1998; 98US-0088701P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Breton G, Bush D;

XX WPI; 2003-576092/54.

DR N-PSDB; ADA32476.

XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents

PT for diagnosing a bacterial disease, as components of antibacterial

PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for

PT plants.

XX Example; SEQ ID NO 7889; 328pp; English.

XX The invention relates to isolated Acinetobacter baumannii nucleic acids.

XX The A. baumannii nucleic acids and polypeptides are useful as reagents

CC for diagnosing a bacterial disease, as components of antibacterial

CC vaccines, as targets for antibacterial drugs, to detect the presence of

CC A. baumannii and other Acinetobacter species in a sample, in screening

CC compounds for the ability to interfere with the A. baumannii life cycle

CC or to inhibit A. baumannii infection, and as biocontrol agents for

CC plants. The present sequence represents the amino acid sequence of an A.

CC baumannii protein.

XX SQ Sequence 269 AA;

Alignment Scores:

Pred. No.: 22 Length: 269

Score: 9.00 Matches: 9

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.31% Indels: 0

DB: Gaps: 0

US-09-729-264-1 (1-1175) x ADA36602 (1-269)

QY 1057 GTAGCTGTGGCCCTCTCACCAGCGG 1083

DB 17 ValAlaValAlaLeuLeuThrSerGly 25

RESULT 28

AAR22653

ID AAR22653 standard; protein; 331 AA.

XX AC AAR22653;

XX AC AAR22653;

DT 25-MAR-2003 (revised)

DT 09-OCT-1992 (first entry)

XX AC

XX Transglutaminase (expressed in E. coli).

XX BTG; acyl rearrangement; deamination; yeast; actinomycetes.

XX Synthetic.

XX EP481504-A.

XX 22-APR-1992.

XX 18-OCT-1991; 91EP-00117813.

XX 19-OCT-1990; 90JP-00282566.

XX (AMANO) AMANO PHARM KK.

XX (AJIN) AJINOMOTO KK.

XX Takagi H, Arafuka S, Matsui H, Washizu K, Ando K, Koikeda S;

XX WPI; 1992-133808/17.

XX N-PSDB; AAQ24207.

XX DNA fragment encoding transglutaminase - is inserted into vector, e.g.

XX PnJ1053-BTG, for protein expression.

XX Disclosure; Page 3; 55pp; English.

XX The protein sequence given has transglutaminase (BTG) activity. The DNA

CC encoding this protein has a base sequence which can be used suitably in

CC an expression system using E. coli or yeast as a host. The base sequence

CC can be compared to those given in AAQ24197 and AAQ24200 which are derived

CC from Actinomycetes sp. BTG catalyses an acyl rearrangement reaction of a

CC gamma-carboxyamide gp. of glutamine. It introduces intra- or

CC intermolecular formation of epsilon-(gamma-Gln)-Lys cross-linking when an

CC epsilon-amino gp. of a Lys residue acts as an acyl receptor. When water

CC acts as an acyl acceptor the enzyme accelerates the conversion of Gln

CC residues to Glu residues by deamination. The enzyme is used in the prodn.

CC of gelled foods, gelled cosmetics, yogurt, gelatin, cheese etc. It is

CC also used in the prodn. of thermally stable materials such as

CC microcapsules and carriers of immobilized enzymes. The DNA sequence given

CC allows the prodn. of BTG efficiently and in large quantity. (Updated on

CC 25-MAR-2003 to correct PA field.)

XX SQ Sequence 331 AA;

Alignment Scores:

Pred. No.: 21.6 Length: 331

Score: 9.00 Matches: 9

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.31% Indels: 0

DB: Gaps: 0

US-09-729-264-1 (1-1175) x AAR22653 (1-331)

QY 356 CCTTACCGTCCAGTTATGGAGAGCT 382

DB 19 ProTyrArgProSerTyrGlyArgAla 27

RESULT 29

AAY33665

ID AAY33665 standard; protein; 331 AA.

XX AC AAY33665;

AC AAY33665;

XX 17-OCT-2003 (revised)  
DT 07-JAN-2000 (first entry)  
XX  
XX Streptovorticillium mobaraense transglutaminase protein fragment.  
DE  
XX Transglutaminase; food industry; pharmaceutical industry; texture;  
KW cosmetic industry; proteinaceous material; gel strength; viscosity;  
KW breaking strength; elasticity; taste.  
XX  
XX Streptomyces mobaraensis.  
OS  
XX DE19814860-A1.  
PN  
XX 07-OCT-1999.  
PD  
XX 02-APR-1998; 98DE-01014860.  
PF  
XX 02-APR-1998; 98DE-01014860.  
PR  
XX (FUCH/) FUCHSBAUER H.  
PA  
XX Fuchsbaumer H, Pasternack R, Dorsch S, Otterbach J, Robenek I;  
PI Mainusch M, Dauscher C;  
PI  
XX WPI; 1999-552288/47.  
DR  
XX Bacterial transglutaminase polypeptides useful for polymerizing proteins,  
PT e.g. to modify the properties of food, pharmaceutical or cosmetic  
PT products.  
XX  
XX Claim 12; Page 25-26; 44pp; German.  
PS  
XX This invention describes a novel bacterial transglutaminase polypeptide.  
CC (I) can be used, e.g. in the food, pharmaceutical and cosmetic  
CC industries, to polymerize proteinaceous materials in order to improve  
CC their properties, e.g. texture, gel strength, breaking strength,  
CC viscosity, elasticity or taste. (I) can also be used to immobilize  
CC enzymes and antibodies. This sequence represents a transglutaminase  
CC protein fragment isolated from Streptovorticillium mobaraense. (Updated  
CC on 17-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 331 AA;  
Alignment Scores:  
Pred. No.: 21.6 Length: 331  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 2 Gaps: 0  
US-09-729-264-1 (1-1175) x AAY33665 (1-331)  
QY 356 CCTTACCGTCCAGTTATGGAGAGCT 382  
Db 19 ProTyrArgProSerTyrGlyArgAla 27  
RESULT 30  
AAY33662  
ID AAY33662 standard; protein; 331 AA.  
XX  
AC AAY33662;  
XX  
DT 17-OCT-2003 (revised)  
DT 07-JAN-2000 (first entry)  
XX  
XX Streptovorticillium S-8112 transglutaminase protein fragment.  
DE  
XX Transglutaminase; food industry; pharmaceutical industry; texture;  
KW cosmetic industry; proteinaceous material; gel strength; viscosity;  
KW breaking strength; elasticity; taste.  
XX

OS Streptomyces sp.  
XX DE19814860-A1.  
PN  
XX 07-OCT-1999.  
PD  
XX 02-APR-1998; 98DE-01014860.  
PF  
XX 02-APR-1998; 98DE-01014860.  
PR  
XX (FUCH/) FUCHSBAUER H.  
PA  
XX Fuchsbaumer H, Pasternack R, Dorsch S, Otterbach J, Robenek I;  
PI Mainusch M, Dauscher C;  
PI  
XX WPI; 1999-552288/47.  
DR  
XX Bacterial transglutaminase polypeptides useful for polymerizing proteins,  
PT e.g. to modify the properties of food, pharmaceutical or cosmetic  
PT products.  
XX  
XX Claim 1; Page 23-24; 44pp; German.  
PS  
XX This invention describes a novel bacterial transglutaminase polypeptide.  
CC (I) can be used, e.g. in the food, pharmaceutical and cosmetic  
CC industries, to polymerize proteinaceous materials in order to improve  
CC their properties, e.g. texture, gel strength, breaking strength,  
CC viscosity, elasticity or taste. (I) can also be used to immobilize  
CC enzymes and antibodies. This sequence represents a transglutaminase  
CC isolated from Streptovorticillium S-8112. (Updated on 17-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 331 AA;  
Alignment Scores:  
Pred. No.: 21.6 Length: 331  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 2 Gaps: 0  
US-09-729-264-1 (1-1175) x AAY33662 (1-331)  
QY 356 CCTTACCGTCCAGTTATGGAGAGCT 382  
Db 19 ProTyrArgProSerTyrGlyArgAla 27  
RESULT 31  
AAY67770  
ID AAY67770 standard; protein; 331 AA.  
XX  
AC AAY67770;  
XX  
DT 17-OCT-2003 (revised)  
DT 01-APR-1999 (first entry)  
XX  
XX A transglutaminase enzyme sequence.  
DE  
XX Transglutaminase; microbial; gelled food; jelly; yogurt; cheese;  
KW cosmetic; meat quality; microcapsule production; high thermal stability;  
KW carrier; immobilised enzyme.  
XX  
XX Streptomyces sp.  
OS  
XX EP889133-A2.  
FN  
XX 07-JAN-1999.  
PD  
XX 02-JUL-1998; 98EP-00112315.  
PF  
XX 04-JUL-1997; 97JP-00180010.  
PR



XX Claim 2; Page 100-102; 126pp; Japanese.

XX The present invention describes a method for designing and constructing a

CC variant transglutaminase by estimating the binding site of

CC Streptococcus mobaraensis with N-terminal aspartic acid deleted -

CC substrate based on the stereo-structure obtained by X-ray analysis of the

CC crystalline structure of MTG crystals, and e.g. substituting amino acid

CC residues located at the substrate-binding site of the transglutaminase.

CC The method can be used for designing and constructing a variant

CC transglutaminase. The obtained transglutaminases can be used in food

CC processing. The modified transglutaminases have improved transglutaminase

CC activity and thermal stability, substrate-specificity and an less

CC required optimum pH. The present sequence represents a transglutaminase

CC isolated from Streptococcus mobaraensis (also called Streptomyces

CC mobaraensis)

XX SQ Sequence 331 AA;

Alignment Scores:

Pred. No.:	21.6	Length:	331
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.31%	Indels:	0
DB:	5	Gaps:	0

US-09-729-264-1 (1-1175) x ABB06742 (1-331)

QY 356 CCTTACCGTCCAAAGTTATGGGAGGCT 382

DB 19 ProTyrArgProSerTyrGlyArgala 27

RESULT 34

AAW67771

ID AAW67771 standard; protein; 332 AA.

AC AAW67771;

XX 01-APR-1999 (first entry)

DE Protein encoded by high expression transglutaminase gene.

XX Transglutaminase; microbial; gelled food; jelly; yogurt; cheese;

KW cosmetic; meat quality; microcapsule production; high thermal stability;

KW carrier; immobilised enzyme.

XX Synthetic.

OS Streptomyces sp.

XX EP889133-A2.

XX 07-JAN-1999.

XX 02-JUL-1998; 98EP-00112315.

XX 04-JUL-1997; 97JP-00180010.

XX (AJIN ) AJINOMOTO CO INC.

PA Yokoyama K, Nakamura N, Miwa T, Seguro K;

PI WPI; 1999-062664/06.

DR N-PSDB; AAW81508.

XX New microbial transglutaminase with N-terminal aspartic acid deleted -

PT allowing high level recombinant production without added methionine in E.

PT coli, useful in production of gelled foods, cosmetics etc.

XX Example 1; Page 18-23; 56pp; English.

PS The present sequence is encoded by the high expression transglutaminase

XX gene present in plasmid pTRPMTG-02. The gene is derived from

CC Streptococcus sp., and is codon altered, using oligonucleotides

CC AAV81521-60, for expression in Escherichia coli. The specification

CC describes a new microbial transglutaminase that has the N-terminal

CC aspartic acid of transglutaminase deleted. Eliminating the N-terminal Asp

CC from microbial transglutaminase allows efficient removal of the terminal

CC Met residue added when the protein is expressed in E. coli. The E. coli

CC methionine aminopeptidase acts well on Met-Ser but only poorly on Met-

CC Asp, so problems of antigenicity associated with Met-terminated proteins

CC are avoided. Recombinant transglutaminase is used to produce gelled foods

CC (jellies, yogurt and cheeses) or cosmetics, to improve the quality of

CC meat, in the production of materials for microcapsules of high thermal

CC stability and as a carrier for immobilised enzymes

XX SQ Sequence 332 AA;

Alignment Scores:

Pred. No.:	21.5	Length:	332
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.31%	Indels:	0
DB:	2	Gaps:	0

US-09-729-264-1 (1-1175) x AAW67771 (1-332)

QY 356 CCTTACCGTCCAAAGTTATGGGAGGCT 382

DB 20 ProTyrArgProSerTyrGlyArgala 28

RESULT 35

AAW12809

ID AAW12809 standard; protein; 332 AA.

XX AAW12809;

XX 24-NOV-2000 (first entry)

DE Transglutaminase protein sequence SEQ ID NO:1.

XX Transglutaminase; gelled food; jelly; yoghurt; cheese.

XX Unidentified.

XX WO200040706-A1.

XX 13-JUL-2000.

XX 24-DEC-1999; 99WO-JP007250.

XX 28-DEC-1998; 98JP-00373131.

XX (AJIN ) AJINOMOTO CO INC.

XX Yokoyama K, Ono K, Ejima D;

XX WPI; 2000-475826/41.

XX N-PSDB; AAA73025.

XX Production of active transglutaminase from denatured enzyme by two-stage

PT refolding process for industrial production of active enzyme for use in

PT food production.

XX Disclosure; Page 48-50; 74pp; Japanese.

XX The present invention describes a method for producing active

CC transglutaminase from denatured enzyme. The method comprises: (i) forming

CC an intermediate structure of the enzyme having transglutaminase activity

CC under acidic conditions in an aqueous medium; and (ii) forming a higher-

CC level structure of the enzyme having transglutaminase activity under

CC neutral conditions in an aqueous medium. The method can be used for

CC industrial production of active transglutaminase from denatured material

CC (such as recombinant transglutaminase) which can be used in the food

CC industry for the production of gelled foods such as jellies, yoghurts and

CC cheeses, and for the production of gelled cosmetics. The present sequence  
CC represents a transglutaminase which is used in the exemplification from  
CC the present invention

Sequence 332 AA;

CC processing. The modified transglutaminases have improved transglutaminase  
CC activity and thermal stability, substrate-specificity and an less  
CC required optimum pH. The present sequence represents a transglutaminase  
CC isolated from *Streptovirgillum mobaraense* (also called *Streptomyces*  
CC *mobaraensis*)

CC cheeses, and for the production of gelled cosmetics. The present sequence  
CC represents a transglutaminase which is used in the exemplification from  
CC the present invention

Sequence 332 AA;



CC crosslinking solution with a biological agent (I); and (ii) incubating so  
 CC that Ag is converted to crosslinked products. The pAg has virucide,  
 CC antibacterial, fungicide, antiparasitic, cytostatic, nootropic and  
 CC neuroprotective activities, and can be used in vaccines. The method is  
 CC useful in preparing immunogenic compositions, using disease-specific  
 CC compounds optionally modified to include a (i)-reactive amino acid that  
 CC are combined in crosslinking solution then treated with (i), that can be  
 CC used as vaccines. pAg, and other related antigenic compositions, are  
 CC useful in human and veterinary medicine, particularly as vaccines, for  
 CC treatment and prevention of infections (viral, bacterial, fungal or  
 CC parasitic), cancers and Alzheimer's disease, also of immune-related or  
 CC autoimmune diseases. Antibodies against pAg are useful as diagnostic  
 CC reagents and crosslinked proteins can also be used industrially, e.g. in  
 CC food or leather processing, in cosmetics and as enzyme carriers. The  
 CC present sequence represents a Streptomyces mobaraensis transglutaminase  
 CC (EC 2.3.2.13), which is used in an example from the present invention.

XX SQ Sequence 332 AA;

Alignment Scores:  
 Pred. No.: 21.5 Length: 332  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.31% Indels: 0  
 DB: 7 Gaps: 0

US-09-729-264-1 (1-1175) x ADB37643 (1-332)

QY 356 CCTTACCGTCACAGTTATGGAGAGCT 382

DB 20 ProTyrArgProSerTyrGlyArgala 28

RESULT 38

AAR49048

ID AAR49048 standard; protein; 346 AA.

AC AAR49048;

XX 10-MAR-2003 (revised)

DT 20-SEP-1994 (first entry)

XX Bacterial transglutaminase.

XX Bacterial transglutaminase; BTG; expression; active; inactive;  
 inclusion body.

XX Escherichia coli.

OS Synthetic.

XX JF06030771-A.

XX 08-FEB-1994.

XX 14-JUL-1992; 92JP-00187038.

XX 14-JUL-1992; 92JP-00187038.

XX (AJIN ) AJINOMOTO KK.

XX WPI; 1994-079294/10.

DR N-PSDB; AAQ55983.

XX Prepn. of bacterial trans-glutaminase in large amts. - by expression of  
 fusion protein in E. coli bacterial trans-glutaminase.

PS Claim 1; Page 7-8; 13pp; Japanese.

XX A fused protein contains amino acids 16-346 of BTG (AAQ55983) and a

CC hydrophilic peptide at the amino terminal. Expression of DNA encoding

CC this protein in E. coli allow large scale prodn. of BTG. An active BTG

CC can be prepd. from the inactive fused protein inclusion body. (Updated on

CC 10-MAR-2003 to add missing OS field.)

XX SQ Sequence 346 AA;

Alignment Scores:

Pred. No.: 21.5 Length: 346  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.31% Indels: 0  
 DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x AAR49048 (1-346)

QY 356 CCTTACCGTCACAGTTATGGAGAGCT 382

DB 34 ProTyrArgProSerTyrGlyArgala 42

RESULT 39

ADB37647

ID ADB37647 standard; protein; 355 AA.

AC ADB37647;

XX 04-DEC-2003 (first entry)

XX Streptomyces mobaraensis 6X-His-TCase fusion protein SEQ ID NO:10.

XX polyvalent antigen; pAg; antigen; Ag; crosslinking; biological agent;  
 virucide; antibacterial; fungicide; antiparasitic; cytostatic; nootropic;  
 neuroprotective; vaccine; immunogenic; antigenic; medicine; infection;  
 cancer; Alzheimer's disease; immune-related disease; autoimmune disease;  
 Streptomyces mobaraensis; transglutaminase; EC 2.3.2.13; enzyme;  
 fusion protein.

XX Synthetic.

OS Streptomyces mobaraensis.

XX Key Location/Qualifiers

FT Misc-difference 300 /note= "encoded by AT"

XX WO2003074004-A2.

XX 12-SEP-2003.

XX 03-MAR-2003; 2003WO-US006661.

XX 01-MAR-2002; 2002US-0361166P.

XX 08-MAR-2002; 2002US-0363445P.

XX 28-AUG-2002; 2002US-00231063.

XX 28-AUG-2002; 2002US-00231114.

XX 28-AUG-2002; 2002US-00231213.

XX 28-AUG-2002; 2002US-00231298.

XX 28-AUG-2002; 2002US-00231470.

XX (CHOU/) CHOU S.

XX Chou S;

XX WPI; 2003-756754/71.

DR N-PSDB; ADB37646.

XX Preparation of polyvalent antigen, useful in vaccines, comprises  
 crosslinking antigen in presence of biological agent, especially  
 transglutaminase, and derived antibodies.

XX Claim 69; Page 128; 130pp; English.

XX The present invention describes a method for producing a polyvalent  
 antigen (pAg). The method comprises: (i) treating an antigen (Ag) in a  
 CC crosslinking solution with a biological agent (I); and (ii) incubating so  
 CC that Ag is converted to crosslinked products. The pAg has virucide,  
 CC antibacterial, fungicide, antiparasitic, cytostatic, nootropic and

CC neuroprotective activities, and can be used in vaccines. The method is  
CC useful in preparing immunogenic compositions, using disease-specific  
CC compounds optionally modified to include a (I)-reactive amino acid that  
CC are combined in crosslinking solution then treated with (I), that can be  
CC used as vaccines. pAg, and other related antigenic compositions, are  
CC useful in human and veterinary medicine, particularly as vaccines, for  
CC treatment and prevention of infections (viral, bacterial, fungal or  
CC parasitic), cancers and Alzheimer's disease, also of immune-related or  
CC autoimmune diseases. Antibodies against pAg are useful as diagnostic  
CC reagents and crosslinked proteins can also be used industrially, e.g. in  
CC food or leather processing, in cosmetics and as enzyme carriers. The  
CC present sequence represents a Streptomyces mobaraensis recombinant 6X-His  
CC -transglutaminase (EC 2.3.2.13) fusion protein, which is used in an  
CC example from the present invention.

XX Sequence 355 AA;

Alignment Scores:                      Length:                      355  
Pred. No.:                      21.4                      Matches:                      9  
Score:                      9.00                      Conservative:                      0  
Percent Similarity:                      100.00%                      Mismatches:                      0  
Best Local Similarity:                      100.00%                      Indels:                      0  
Query Match:                      2.31%                      Gaps:                      0  
DB:                      7

US-09-729-264-1 (1-1175) x ADB37647 (1-355)

QY                      356 CTTACGTCACAGTTATGGAGAGCT 382

Db                      43 ProTyArgProSerTyRGlyArgAla 51

RESULT 40

ABG22212  
ID                      ABG22212 standard; protein; 360 AA.

AC                      ABG22212;

DT                      18-FEB-2002 (first entry)

XX                      Novel human diagnostic protein #22203.

XX                      Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW                      food supplement; medical imaging; diagnostic; genetic disorder.

XX                      Homo sapiens.

OS                      WO200175067-A2.

XX                      11-OCT-2001.

XX                      30-MAR-2001; 2001WO-US008631.

XX                      31-MAR-2000; 2000US-00540217.

PR                      23-AUG-2000; 2000US-00649167.

XX                      (HYSE-) HYSEQ INC.

XX                      Drmanac RT, Liu C, Tang YT;

XX                      WPI; 2001-639362/73.

DR                      N-PSDB; AAS86399.

XX                      New isolated polynucleotide and encoded polypeptides, useful in  
PT                      diagnostics, forensics, gene mapping, identification of mutations  
PT                      responsible for genetic disorders or other traits and to assess  
PT                      biodiversity.

XX                      Claim 20; SEQ ID NO 52571; 103pp; English.

XX                      The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC                      sequences. (I) is useful as hybridisation probes, polymerase chain  
CC                      reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC                      and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activities. The  
CC polypeptide and polynucleotide sequences have application in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 360 AA;

Alignment Scores:                      Length:                      360  
Pred. No.:                      21.4                      Matches:                      9  
Score:                      9.00                      Conservative:                      0  
Percent Similarity:                      100.00%                      Mismatches:                      0  
Best Local Similarity:                      100.00%                      Indels:                      0  
Query Match:                      2.31%                      Gaps:                      0  
DB:                      4

US-09-729-264-1 (1-1175) x ABG22212 (1-360)

QY                      476 CTGGAGCTCGGTCTCTCTGTGTCAGCCA 502

Db                      189 LeuGlyAlaArgSerProGlyGlnPro 197

RESULT 41

AAR22651  
ID                      AAR22651 standard; protein; 406 AA.

AC                      AAR22651;

XX                      27-AUG-2003 (revised)

DT                      25-MAR-2003 (revised)

DT                      09-OCT-1992 (first entry)

XX                      Transglutaminase.

XX                      BTG; acyl rearrangement; deamination.

XX                      Fungi.

XX                      Key                      Location/Qualifiers

FT                      Peptide                      1..75

FT                      Protein                      /label= sig\_peptide

FT                                           76..406

FT                                           /label= mat\_BTG

XX                      EP481504-A.

XX                      22-APR-1992.

XX                      18-OCT-1991; 91EP-00117813.

XX                      19-OCT-1990; 90JP-00282566.

XX                      (AMANO) AMANO PHARM KK.

XX                      (AJIN) AJINOMOTO KK.

XX                      Takagi H, Arafuka S, Matsui H, Washizu K, Ando K, Koikeda S;

XX                      WPI; 1992-133808/17.

DR                      N-PSDB; AAQ24197, AAQ24201.

XX                      DNA fragment encoding trans:glutaminase - is inserted into vector, e.g.

PT

PT PnJ1053-BTG, for protein expression.  
 XX  
 PS Disclosure; Page 42; 55pp; English.  
 XX  
 CC The mature transglutaminase enzyme (BTG) can be derived from two  
 CC different species, Streptococcus and Actinomyces. It catalyzes an  
 CC acyl rearrangement reaction of a gamma-carboxamide gp. of glutamine. It  
 CC introduces intra- or intermolecular formation of epsilon-(gamma-Gln)-Lys  
 CC cross-linking when an epsilon-amino gp. of a Lys residue acts as an acyl  
 CC receptor. When water acts as an acyl acceptor the enzyme accelerates the  
 CC conversion of Gln residues to Glu residues by deamination. The enzyme is  
 CC used in the prodn. of gelled foods, gelled cosmetics, yogurt, gelatin,  
 CC cheese etc. It is also used in the prodn. of thermally stable materials  
 CC such as microcapsules and carriers of immobilized enzymes. The DNA  
 CC sequence given allows the prodn. of BTG efficiently and in large  
 CC quantity. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 27-  
 CC AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 406 AA;  
 Alignment Scores:  
 Pred. No.: 21.1 Length: 406  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.31% Indels: 0  
 DB: 2 Gaps: 0  
 US-09-729-264-1 (1-1175) x AAR22651 (1-406)  
 QY 356 CCTTACCGTCCAAGTTATGGAGAGCT 382  
 Db 94 ProTyrArgProSerTyrGlyArgAla 102  
 RESULT 42  
 AAB97831  
 ID AAB97831 standard; protein; 407 AA.  
 AC AAB97831;  
 XX  
 DT 06-AUG-2003 (revised)  
 DT 09-AUG-2001 (first entry)  
 XX  
 DE S. mobaraense IFO 13819 transglutaminase protein SEQ ID NO:4.  
 XX  
 KW Streptococcus cinnamomeum IFO 12852; Streptomyces; actinomycete;  
 KW Streptococcus mobaraense IFO 13819; transglutaminase.  
 XX  
 OS Streptomyces mobaraensis.  
 XX  
 PN WO200129187-A1.  
 XX  
 PD 26-APR-2001.  
 XX  
 PF 13-OCT-2000; 2000WO-JP007135.  
 XX  
 PR 18-OCT-1999; 99JP-00295649.  
 XX  
 PA (AJIN ) AJINOMOTO CO INC.  
 XX  
 PI Taguchi S, Momose H;  
 XX  
 DR WPI; 2001-300330/31.  
 DR N-PSDB; AAB20188.  
 XX  
 PT Streptomyces sp. carrying an actinomycete-derived gene and promoter for  
 PT producing high yields of transglutaminase.  
 XX  
 PS Disclosure; Page 33-36; 41pp; Japanese.  
 XX  
 CC The present invention describes a Streptomyces sp. containing a gene  
 CC construct comprising actinomycete-derived transglutaminase gene and  
 CC promoter. Also described are methods for producing pro-transglutaminase

CC and active transglutaminase. The gene construct can be used in the  
 CC production of large amounts of transglutaminase. The present sequence  
 CC represents Streptococcus mobaraense IFO 13819 transglutaminase,  
 CC which is given in the exemplification of the present invention. (Updated  
 CC on 06-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 407 AA;  
 Alignment Scores:  
 Pred. No.: 21.1 Length: 407  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.31% Indels: 0  
 DB: 4 Gaps: 0  
 US-09-729-264-1 (1-1175) x AAB97831 (1-407)  
 QY 356 CCTTACCGTCCAAGTTATGGAGAGCT 382  
 Db 95 ProTyrArgProSerTyrGlyArgAla 103  
 RESULT 43  
 AAB81166  
 ID AAB81166 standard; protein; 407 AA.  
 XX  
 AC AAB81166;  
 XX  
 DT 11-SEP-2003 (revised)  
 DT 13-JUL-2001 (first entry)  
 XX  
 DE Prepro-transglutaminase amino acid sequence.  
 XX  
 KW Corynebacterium mobaraense; transglutaminase; food processing.  
 XX  
 OS Streptomyces mobaraensis.  
 XX  
 PN WO200123591-A1.  
 XX  
 PD 05-APR-2001.  
 XX  
 PF 29-SEP-2000; 2000WO-JP006780.  
 XX  
 PR 30-SEP-1999; 99JP-00280098.  
 PR 28-JUN-2000; 2000JP-00194043.  
 XX  
 PA (AJIN ) AJINOMOTO CO INC.  
 XX  
 PI Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Matsui H;  
 XX  
 DR WPI; 2001-266172/27.  
 DR N-PSDB; AAF86283.  
 XX  
 PT Efficient secretory production of foreign proteins e.g. transglutaminase  
 PT employing transformant corynebacterium, simply on industrial scale  
 PT with direct recovery for use in food processing and pharmaceutical  
 PT industry.  
 XX  
 PS Example 1; Page 90-93; 151pp; Japanese.  
 XX  
 CC This invention relates to a process for the production of a foreign  
 CC secretory protein through the construction of a recombinant corynebacterium.  
 CC The corynebacterium is transformed with an expression  
 CC construct in which DNA encoding a target foreign protein pro-sequence is  
 CC ligated to the downstream region of DNA encoding the signal peptide  
 CC domain of a corynebacterium protein. Following transformation with  
 CC the vector, the bacterium is cultured, and the pro-peptide cleaved from  
 CC the expressed protein. Transglutaminases produced using this process are  
 CC useful in the food processing and pharmaceutical industries. The present  
 CC sequence prepro-transglutaminase. The protein is used in an example  
 CC illustrating the method of the invention. (Updated on 11-SEP-2003 to  
 CC standardise OS field)  
 XX

SQ Sequence 407 AA;

Alignment Scores:  
 Pred. No.: 21.1 Length: 407  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.31% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x AAB81166 (1-407)

QY 356 CCTTACCGTCCAGTTATGGGAGAGCT 382  
 |||||  
 Db 95 ProTyrArgProSerTyrGlyArgAla 103

RESULT 44  
 ABU07390  
 ID ABU07390 standard; protein; 407 AA.  
 XX AC ABU07390;  
 XX DT 23-OCT-2003 (revised)  
 XX DT 28-JAN-2003 (first entry)  
 XX Foreign protein #1.  
 XX DE Coryneform bacterium; signal peptide domain; food processing; medicine;  
 XX KW cosmetic; transglutaminase; human epithelial growth factor.  
 XX KW Streptomyces mobaraensis.  
 XX OS WO200281694-A1.  
 XX PN 17-OCT-2002.  
 XX PD 27-MAR-2002; 2002WO-JP002978.  
 XX PF 30-MAR-2001; 2001JP-00098808.  
 XX PR (AJIN ) AJINOMOTO CO INC.  
 XX PA Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Heima H, Matsui H;  
 XX PI WPI; 2003-058550/05.  
 XX DR N-PSDB; ABX10285.  
 XX Secretion production of foreign proteins by culturing transformant  
 PT coryneform bacteria, applicable in producing e.g. industrially-useful  
 PT transglutaminase and human epithelial growth factor.  
 XX PS Disclosure; Page 68-71; 117pp; Japanese.

The invention relates to a method for producing a foreign protein by  
 CC culturing a mutant strain of a coryneform bacterium that contains an  
 CC expression gene construct to secrete the foreign protein, followed by  
 CC recovery of the produced foreign protein. The gene construct is obtained  
 CC by ligation of a nucleic acid encoding a signal peptide domain  
 CC originating from a coryneform bacterium, to downstream of a functioning  
 CC promoter sequence in the coryneform bacterium and also by ligation of a  
 CC nucleic acid encoding a signal peptide, to downstream of a nucleic acid  
 CC sequence encoding the signal peptide. The method is useful for the  
 CC production of a foreign protein which is applicable in producing e.g.  
 CC industrially-useful transglutaminase and human epithelial growth factor  
 CC for use in medicine, cosmetics and food processing. This sequence  
 CC represents a foreign protein of the invention. (Updated on 23-OCT-2003 to  
 CC standardise OS field)

SQ Sequence 407 AA;

Alignment Scores:  
 Pred. No.: 21.1 Length: 407  
 Score: 9.00 Matches: 9

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.31% Indels: 0  
 DB: 6 Gaps: 0

US-09-729-264-1 (1-1175) x ABU07390 (1-407)

QY 356 CCTTACCGTCCAGTTATGGGAGAGCT 382  
 |||||  
 Db 95 ProTyrArgProSerTyrGlyArgAla 103

RESULT 45  
 AAR49049  
 ID AAR49049 standard; protein; 408 AA.  
 XX AC AAR49049;  
 XX DT 10-MAR-2003 (revised)  
 XX DT 20-SEP-1994 (first entry)  
 XX BTG-contg. sequence.  
 XX DE Bacterial transglutaminase; BTG; expression; active; inactive;  
 XX KW inclusion body.  
 XX KW Escherichia coli.  
 XX OS Synthetic.  
 XX PN JP06030771-A.  
 XX PD 08-FEB-1994.  
 XX PF 14-JUL-1992; 92JP-00187038.  
 XX PR 14-JUL-1992; 92JP-00187038.  
 XX PA (AJIN ) AJINOMOTO KK.  
 XX DR WPI; 1994-079294/10.  
 XX DR N-PSDB; AAQ55984.  
 XX PT Prepn. of bacterial trans-glutaminase in large amts. - by expression of  
 PT fusion protein in E. coli bacterial trans-glutaminase.  
 XX PS Disclosure; Page 8-10; 13pp; Japanese.

A fused protein contains amino acids 16-346 of BTG (AAQ55983) and a  
 CC hydrophilic peptide at the amino terminal. Expression of DNA encoding  
 CC this protein in E. coli allow large scale prodn. of BTG. An active BTG  
 CC can be prepd. from the inactive fused protein inclusion body. (Updated on  
 CC 10-MAR-2003 to add missing OS field.)

SQ Sequence 408 AA;

Alignment Scores:  
 Pred. No.: 21.1 Length: 408  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.31% Indels: 0  
 DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x AAR49049 (1-408)

QY 356 CCTTACCGTCCAGTTATGGGAGAGCT 382  
 |||||  
 Db 96 ProTyrArgProSerTyrGlyArgAla 104

Search completed: September 18, 2004, 22:53:09  
 Job time : 74.4053 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 18, 2004, 22:47:02 ; Search time 13.9373 Seconds  
(without alignment)  
8704.746 Million cell updates/sec

Title: us-09-729-264-1

Perfect score: 389

Sequence: 1 ctgtctgcccattcgataa.....gtaatacaactgttagtatag 1175

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Word size: 0

Total number of hits satisfying chosen parameters: 400536

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Listing first 135 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-O=/cgn2\_1/uspto.spool p/US09729264/runat 17092004 155108 2401/app query.fasta\_1.4117  
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=oli.ra1 -MINMATCH=0.1 -IOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=135 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=0 -ALIGN=45 -MODE=LOCAL  
-USER=US09729264 @CGN 1 1 74 @runat 17092004 155108 2401 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DRV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	2.3	269	4	US-09-328-352-7889
2	9	2.3	331	1	US-08-136-993-1
3	9	2.3	331	3	US-09-109-063-1
4	9	2.3	331	3	US-08-793-426A-3
5	9	2.3	331	3	US-09-294-565-3
6	9	2.3	331	4	US-09-448-310-1
7	9	2.3	406	1	US-08-136-993-13
8	9	2.3	410	4	US-10-022-809A-2
9	9	2.3	410	4	US-10-022-809A-5
10	9	2.4	605	4	US-09-976-594-616
11	9	2.4	908	4	US-08-714-741-44
12	8	2.1	81	4	US-09-621-976-5662

Sequence 301, App	US-09-188-930-301	82	3	2.1	8	13
Sequence 301, App	US-09-312-283C-301	82	4	2.1	8	14
Sequence 29946, A	US-09-252-991A-29946	202	4	2.1	8	15
Sequence 28250, A	US-09-252-991A-28250	249	4	2.1	8	16
Sequence 9994, Ap	US-09-489-039A-9994	271	4	2.1	8	17
Sequence 33090, A	US-09-252-991A-33090	460	4	2.1	8	18
Sequence 24229, A	US-09-252-991A-24229	495	4	2.1	8	19
Sequence 3, Appli	US-08-891-298-3	606	4	2.1	8	20
Sequence 11, Appli	US-08-653-648A-11	606	4	2.1	8	21
Sequence 10, Appli	US-09-564-418-10	606	4	2.1	8	22
Sequence 4394, Ap	US-09-134-001C-4394	629	4	2.1	8	23
Sequence 37, Appli	US-09-644-460-37	717	4	2.1	8	24
Sequence 18, Appli	US-09-035-648-18	747	3	2.1	8	25
Sequence 18, Appli	US-09-001-951-18	747	3	2.1	8	26
Sequence 18, Appli	US-08-818-829-18	747	4	2.1	8	27
Sequence 27424, A	US-09-252-991A-27424	751	4	2.1	8	28
Sequence 6, Appli	US-07-906-349A-6	801	1	2.1	8	29
Sequence 4643, Ap	US-09-134-000C-4643	808	4	2.1	8	30
Sequence 37, Appli	US-08-630-915A-37	1400	4	2.1	8	31
Sequence 2, Appli	US-09-060-854B-2	1497	4	2.1	8	32
Sequence 7, Appli	US-09-627-650B-7	2508	4	2.1	8	33
Sequence 7, Appli	US-09-436-063C-7	2508	4	2.1	8	34
Sequence 3, Appli	US-09-627-650B-3	2544	4	2.1	8	35
Sequence 3, Appli	US-09-436-063C-3	2544	4	2.1	8	36
Sequence 9, Appli	US-09-627-650B-9	2601	4	2.1	8	37
Sequence 56, Appli	US-08-425-069-56	2601	4	2.1	8	38
Sequence 56, Appli	US-08-317-844B-56	27	1	1.8	7	39
Sequence 390, App	US-08-905-223-390	27	2	1.8	7	40
Sequence 30, Appli	US-09-117-121-30	46	3	1.8	7	41
Sequence 7668, Ap	US-09-328-352-7668	54	4	1.8	7	42
Sequence 15, Appli	US-09-238-303-15	62	4	1.8	7	43
Sequence 15, Appli	US-09-946-239-15	65	3	1.8	7	44
Sequence 5438, Ap	US-09-462-478A-15	65	4	1.8	7	45
Sequence 2862, Ap	US-09-134-000C-5438	68	4	1.8	7	46
Sequence 6091, Ap	US-09-540-236-2862	73	4	1.8	7	47
Sequence 18845, A	US-09-621-976-6091	75	4	1.8	7	48
Sequence 2, Appli	US-09-252-991A-18845	79	4	1.8	7	49
Sequence 2, Appli	US-09-344-529-2	83	4	1.8	7	50
Sequence 278, App	US-07-812-421-2	92	2	1.8	7	51
Sequence 8191, Ap	US-08-905-223-278	98	2	1.8	7	52
Sequence 5460, Ap	US-09-328-352-8191	98	2	1.8	7	53
Sequence 172, App	US-09-621-976-5460	99	4	1.8	7	54
Sequence 6, Appli	US-09-177-249-172	100	4	1.8	7	55
Sequence 5637, Ap	US-09-034-916-6	108	3	1.8	7	56
Sequence 2, Appli	US-09-543-681A-5637	111	3	1.8	7	57
Sequence 2, Appli	US-09-540-014-2	115	4	1.8	7	58
Sequence 26903, A	US-09-540-014-2	122	4	1.8	7	59
Sequence 3, Appli	US-09-252-991A-26903	123	4	1.8	7	60
Sequence 3, Appli	US-08-808-303-3	124	1	1.8	7	61
Sequence 4290, Ap	US-08-996-533-3	124	3	1.8	7	62
Sequence 1, Appli	US-09-107-532A-4290	124	3	1.8	7	63
Sequence 6, Appli	US-08-053-006-1	128	4	1.8	7	64
Sequence 30876, A	US-09-540-014-6	130	1	1.8	7	65
Sequence 17806, A	US-09-252-991A-30876	130	4	1.8	7	66
Sequence 10, Appli	US-09-252-991A-17806	140	4	1.8	7	67
Sequence 10927, A	US-09-554-765-10	146	4	1.8	7	68
Sequence 4212, Ap	US-09-489-039A-10927	147	4	1.8	7	69
Sequence 26692, A	US-09-621-976-4212	154	4	1.8	7	70
Sequence 24956, A	US-09-621-976-4212	154	4	1.8	7	71
Sequence 3906, Ap	US-09-252-991A-24956	154	4	1.8	7	72
Sequence 4792, Ap	US-09-621-976-3906	158	4	1.8	7	73
Sequence 27091, A	US-09-621-976-4792	158	4	1.8	7	74
Sequence 6, Appli	US-09-252-991A-27091	160	2	1.8	7	75
Sequence 8, Appli	US-08-319-704-6	162	1	1.8	7	76
Sequence 12, Appli	US-08-808-303-8	164	1	1.8	7	77
Sequence 12, Appli	US-08-808-303-12	164	1	1.8	7	78
Sequence 12, Appli	US-08-996-533-8	164	3	1.8	7	79
Sequence 1, Appli	US-08-996-533-12	164	3	1.8	7	80
Sequence 28523, A	US-08-801-742-1	164	4	1.8	7	81
Sequence 28858, A	US-09-252-991A-28523	164	4	1.8	7	82
	US-09-252-991A-28858	165	4	1.8	7	83
						84
						85

C 86 7 1.8 172 4 US-09-107-532A-5847  
C 87 7 1.8 176 4 US-09-252-991A-30148  
C 88 7 1.8 176 4 US-09-198-452A-1032  
C 89 7 1.8 180 4 US-09-252-991A-18500  
C 90 7 1.8 183 6 5168049-3  
C 91 7 1.8 186 4 US-09-252-991A-17637  
C 92 7 1.8 192 4 US-09-149-476-419  
C 93 7 1.8 195 4 US-09-071-035-346  
C 94 7 1.8 195 4 US-09-328-352-4743  
C 95 7 1.8 202 4 US-09-134-000C-3660  
C 96 7 1.8 213 4 US-09-252-991A-16621  
C 97 7 1.8 213 4 US-09-252-991A-30140  
C 98 7 1.8 217 1 US-08-621-081A-23  
C 99 7 1.8 221 4 US-09-252-991A-20560  
C 100 7 1.8 222 4 US-09-543-681A-6557  
C 101 7 1.8 223 4 US-09-252-991A-18819  
C 102 7 1.8 223 4 US-09-252-991A-27694  
C 103 7 1.8 225 4 US-09-252-991A-25292  
C 104 7 1.8 227 4 US-09-540-236-2217  
C 105 7 1.8 231 4 US-09-252-991A-26289  
C 106 7 1.8 233 2 US-09-151-611-1  
C 107 7 1.8 233 3 US-09-370-102-1  
C 108 7 1.8 234 3 US-08-924-747-8  
C 109 7 1.8 234 3 US-09-247-373B-8  
C 110 7 1.8 234 3 US-09-296-715-8  
C 111 7 1.8 234 4 US-09-107-532A-3787  
C 112 7 1.8 235 4 US-09-252-991A-19935  
C 113 7 1.8 235 4 US-09-252-991A-25654  
C 114 7 1.8 236 4 US-09-252-991A-18461  
C 115 7 1.8 237 3 US-09-216-001-1  
C 116 7 1.8 237 3 US-08-878-862-1  
C 117 7 1.8 238 3 US-09-234-332-10  
C 118 7 1.8 238 4 US-09-702-705-1813  
C 119 7 1.8 238 4 US-09-736-457-1813  
C 120 7 1.8 238 4 US-09-252-991A-20037  
C 121 7 1.8 238 4 US-09-671-325-1813  
C 122 7 1.8 246 4 US-09-552-225A-12  
C 123 7 1.8 248 4 US-09-252-991A-30100  
C 124 7 1.8 251 4 US-09-252-991A-18685  
C 125 7 1.8 251 4 US-09-489-039A-7987  
C 126 7 1.8 255 4 US-09-134-000C-6320  
C 127 7 1.8 259 4 US-09-252-991A-24558  
C 128 7 1.8 261 4 US-09-134-001C-4193  
C 129 7 1.8 268 4 US-09-554-765-2  
C 130 7 1.8 270 4 US-09-328-352-4859  
C 131 7 1.8 272 2 US-08-492-027A-4  
C 132 7 1.8 273 4 US-09-489-039A-9102  
C 133 7 1.8 274 4 US-09-134-000C-4610  
C 134 7 1.8 276 4 US-09-134-001C-3456  
C 135 7 1.8 277 4 US-09-186-276B-34

## ALIGNMENTS

RESULT 1  
US-09-328-352-7889  
; Sequence 7889, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 7889  
; LENGTH: 269  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-7889  
Alignment Scores:  
Pred. No.: 5.57 Length: 331  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 1 Gaps: 0

Pred. No.: 5.67 Length: 269  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-328-352-7889 (1-269)

QY 1057 GTAGCTGTGGCCCTCTCTACCGCGGG 1083  
|||  
Db 17 ValAlaValAlaLeuThrSerGly 25  
|||

RESULT 2  
US-08-136-993-1  
; Sequence 1, Application US/08136993  
; Patent No. 5420025  
; GENERAL INFORMATION:  
; APPLICANT: Takagi, Hiroshi  
; APPLICANT: Afafuka, Shino  
; APPLICANT: Matsui, Hiroshi  
; APPLICANT: Washizu, Kinya  
; APPLICANT: Ando, Keiichi  
; APPLICANT: Koikeda, Satoshi  
; TITLE OF INVENTION: Recombinant transglutaminase  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue  
; CITY: N.W.  
; STATE: Washington, D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20037-3202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/136,993  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/777,447  
; FILING DATE:  
; APPLICATION NUMBER: JP 2-282566  
; FILING DATE: 19-OCT-1990  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-293-7060  
; TELEFAX: 202-293-7860  
; TELEX: 6491103  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 331 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-136-993-1

Alignment Scores:  
Pred. No.: 5.57 Length: 331  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x US-08-136-993-1 (1-331)

QY 356 CCTTACCGTCCAGTTATGGAGGCT 382  
|||  
Db 19 ProTyArgProSerTyrglyArgAla 27  
|||

```
RESULT 3
US-09-109-063-1
; Sequence 1, Application US/09109063
; Patent No. 6013498
; GENERAL INFORMATION:
; APPLICANT: YOKOYAMA, KEIICHI
; APPLICANT: NAKAMURA, NAMU
; APPLICANT: MIWA, TETSUYA
; APPLICANT: SEGURO, KATSUYA
; TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
; FILE REFERENCE: 0010-0937-0
; CURRENT APPLICATION NUMBER: US/09/109,063
; CURRENT FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: JP 180010/1997
; EARLIER FILING DATE: 1997-07-04
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial
; OTHER INFORMATION: Sequence:TRANSGLUTAMINASE
US-09-109-063-1

Alignment Scores:
Pred. No.:          5.57      Length:      331
Score:              9.00      Matches:      9
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:        2.31%     Indels:      0
DB:                  3       Gaps:        0

US-09-729-264-1 (1-1175) x US-09-109-063-1 (1-331)
QY 356 CCTTACCGTCCAAAGTTATGGGAGAGCT 382
Db 19 ProTyrArgProSerTyrGlyArgAla 27

RESULT 4
US-08-793-426A-3
; Sequence 3, Application US/08793426A
; Patent No. 6100053
; GENERAL INFORMATION:
; APPLICANT: Bech, Lisbeth
; APPLICANT: No. 6100053revang, Iben
; APPLICANT: Halkier, Torben
; APPLICANT: Rasmussen, Grethe
; APPLICANT: Schafer, Thomas
; APPLICANT: Andersen, Jens
; TITLE OF INVENTION: Microbial Transglutaminases, Their
; TITLE OF INVENTION: Production And Use
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6100053o No. 6100053disk of No. 6100053th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,426A
; FILING DATE: 25-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.
; REGISTRATION NUMBER: 36,993
```

```
REFERENCE/DOCKET NUMBER: 4211.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-793-426A-3

Alignment Scores:
Pred. No.:          5.57      Length:      331
Score:              9.00      Matches:      9
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:        2.31%     Indels:      0
DB:                  3       Gaps:        0

US-09-729-264-1 (1-1175) x US-08-793-426A-3 (1-331)
QY 356 CCTTACCGTCCAAAGTTATGGGAGAGCT 382
Db 19 ProTyrArgProSerTyrGlyArgAla 27

RESULT 5
US-09-294-565-3
; Sequence 3, Application US/09294565
; Patent No. 6190879
; GENERAL INFORMATION:
; APPLICANT: Bech, Lisbeth
; APPLICANT: No. 6190879revang, Iben
; APPLICANT: Halkier, Torben
; APPLICANT: Rasmussen, Grethe
; APPLICANT: Schafer, Thomas
; APPLICANT: Andersen, Jens
; TITLE OF INVENTION: Microbial Transglutaminases, Their
; TITLE OF INVENTION: Production And Use
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6190879o No. 6190879disk of No. 6190879th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/294,565
; FILING DATE: 19-APR-1999
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4211.224-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

US-09-294-565-3

Alignment Scores:  
Pred. No.: 5.57 Length: 331  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 3 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-294-565-3 (1-331)

Qy 356 CCTTACCGTCCAAAGTTATGGAGAGCT 382

Db 19 ProTyrArgProSerTyrGlyArgAla 27

RESULT 6

US-09-448-310-1  
; Sequence 1, Application US/09448310  
; Patent No. 6538122  
; GENERAL INFORMATION:  
; APPLICANT: YOKOYAMA, KEIICHI  
; APPLICANT: NAKAMURA, NAMI  
; APPLICANT: MIWA, TETSUYA  
; APPLICANT: SEGURO, KATSUYA  
; TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE  
; FILE REFERENCE: 0010-0937-0  
; CURRENT APPLICATION NUMBER: US/09/448,310  
; CURRENT FILING DATE: 1999-11-24  
; PRIOR APPLICATION NUMBER: 09/109,063  
; PRIOR FILING DATE: 1998-07-02  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 331  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial  
; OTHER INFORMATION: Sequence:TRANSGLUTAMINASE  
US-09-448-310-1

Alignment Scores:  
Pred. No.: 5.57 Length: 331  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-448-310-1 (1-331)

Qy 356 CCTTACCGTCCAAAGTTATGGAGAGCT 382

Db 19 ProTyrArgProSerTyrGlyArgAla 27

RESULT 7

US-08-136-993-13  
; Sequence 13, Application US/08136993  
; Patent No. 5420025  
; GENERAL INFORMATION:  
; APPLICANT: Takagi, Hiroshi  
; APPLICANT: Aratuka, Shino  
; APPLICANT: Matsui, Hiroshi  
; APPLICANT: Washizu, Kinya  
; APPLICANT: Ando, Keiichi  
; APPLICANT: Koikeda, Satoshi  
; TITLE OF INVENTION: Recombinant transglutaminase  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue  
; CITY: N.W.

; STATE: Washington, D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20037-3202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/136,993  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/777,447  
; FILING DATE:  
; APPLICATION NUMBER: JP 2-282566  
; FILING DATE: 19-OCT-1990  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-293-7060  
; TELEFAX: 202-293-7860  
; TELEX: 6491103  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 406 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-136-993-13

Alignment Scores:  
Pred. No.: 5.47 Length: 406  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x US-08-136-993-13 (1-406)

Qy 356 CCTTACCGTCCAAAGTTATGGAGAGCT 382

Db 94 ProTyrArgProSerTyrGlyArgAla 102

RESULT 8

US-10-022-809A-2  
; Sequence 2, Application US/10022809A  
; Patent No. 6660510  
; GENERAL INFORMATION:  
; APPLICANT: LIN, Yi-Shin  
; APPLICANT: LIU, Chang-Hsieh  
; APPLICANT: CHU, Wen-Shen  
; TITLE OF INVENTION: TRANSGLUTAMINASE GENE OF STREPTOVERTICILLUM LADAKANUM AND THE  
; FILE REFERENCE: U 013779-2  
; CURRENT APPLICATION NUMBER: US/10/022,809A  
; CURRENT FILING DATE: 2001-12-17  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patent In version 3.1  
; SEQ ID NO 2  
; LENGTH: 410  
; TYPE: PRT  
; ORGANISM: Streptovorticillum ladakanum  
US-10-022-809A-2

Alignment Scores:  
Pred. No.: 5.46 Length: 410  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-022-809A-2 (1-410)



```
Qy 356 CCTTACCGTCCAAAGTTATGGGAGCT 382
|
Db 98 ProTyrArgProSerTyrGlyArgAla 106

RESULT 9
US-10-022-809A-5
; Sequence 5, Application US/10022809A
; Patent No. 6660510
; GENERAL INFORMATION:
; APPLICANT: LIU, Yi-Shin
; APPLICANT: LIU, Chang-Hsieh
; APPLICANT: CHU, Wen-Shen
; TITLE OF INVENTION: TRANSLUTAMINASE GENE OF STREPTOVERTICILLIUM LADAKANUM AND THE
; TITLE OF INVENTION: TRANSLUTAMINASE ENCODED THEREFROM
; FILE REFERENCE: U 013779-2
; CURRENT APPLICATION NUMBER: US/10/022,809A
; CURRENT FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Streptovorticillium ladakanum
US-10-022-809A-5
Alignment Scores:
Pred. No.: 5.46 Length: 410
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.31% Indels: 0
Gaps: 0
DB:

US-09-729-264-1 (1-1175) x US-10-022-809A-5 (1-410)
Qy 356 CCTTACCGTCCAAAGTTATGGGAGCT 382
|
Db 98 ProTyrArgProSerTyrGlyArgAla 106

RESULT 10
US-09-976-594-616
; Sequence 616, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 616
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1692213CE1
; NAME/KEY: unsure
; LOCATION: 596
; OTHER INFORMATION: unknown or other
US-09-976-594-616
Alignment Scores:
Pred. No.: 5.28 Length: 605
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.37% Indels: 0
Gaps: 0
DB:

US-09-729-264-1 (1-1175) x US-10-022-809A-5 (1-410)
Qy 356 CCTTACCGTCCAAAGTTATGGGAGCT 382
|
Db 98 ProTyrArgProSerTyrGlyArgAla 106

RESULT 11
US-08-714-741-44
; Sequence 44, Application US/08714741
; Patent No. 6500613
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yother, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Tart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF.
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
; TITLE OF INVENTION: PORTIONS AND PRODUCTS
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,741
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer Esq., William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2460
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-714-741-44
Alignment Scores:
Pred. No.: 5.1 Length: 908
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.37% Indels: 0
Gaps: 0
DB:

US-09-729-264-1 (1-1175) x US-08-714-741-44 (1-908)
Qy 833 ACAACGACGGCGGCGAGCAGCAGCG 807
|
Db 321 ThrThrAlaAlaAlaAlaAlaAla 329

RESULT 12
US-09-621-976-5662
```

; Sequence 5662, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 5662  
; LENGTH: 81  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -24...-1  
US-09-621-976-5662

Alignment Scores:  
Pred. No.: 54.5 Length: 81  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-621-976-5662 (1-81)

QY 827 TCGTGTGTGCTGCACTGCTG 850  
Db 12 SerLeuLeuPheLeuGlnLeuLeu 19  
|||||

## RESULT 13

US-09-188-930-301  
; Sequence 301, Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James Greg  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; TITLE OF INVENTION: and Methods For Their Use  
; FILE REFERENCE: 11000.1011c1  
; CURRENT APPLICATION NUMBER: US/09/188,930A  
; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 301  
; LENGTH: 82  
; TYPE: PRT  
; ORGANISM: Mouse  
US-09-188-930-301

Alignment Scores:  
Pred. No.: 54.4 Length: 82  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 3 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-188-930-301 (1-82)

QY 848 CTGCTGCGTGTGTTCTGCTG 871  
Db 35 LeuLeuProLeuLeuPheLeuLeu 42  
|||||

## RESULT 14

US-09-312-283C-301  
; Sequence 301, Application US/09312283C  
; Patent No. 6573095  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James G.  
; APPLICANT: Kumble, Krishanand D.  
; TITLE OF INVENTION: Compositions Isolated from Skin Cells  
; TITLE OF INVENTION: and Methods for Their Use  
; FILE REFERENCE: 11000.1011c2  
; CURRENT APPLICATION NUMBER: US/09/312,283C  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 425  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 301  
; LENGTH: 82  
; TYPE: PRT  
; ORGANISM: Mouse  
US-09-312-283C-301

Alignment Scores:  
Pred. No.: 54.4 Length: 82  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-312-283C-301 (1-82)

QY 848 CTGCTGCGTGTGTTCTGCTG 871  
Db 35 LeuLeuProLeuLeuPheLeuLeu 42  
|||||

## RESULT 15

US-09-252-991A-29946  
; Sequence 29946, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 29946  
; LENGTH: 202  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29946

Alignment Scores:  
Pred. No.: 50.3 Length: 202  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-252-991A-29946 (1-202)

QY 153 GCTTCCAGCCCTGGGAGCGTG 130  
Db 9 AlaSerProGlyArgArgCys 16  
|||||

```

RESULT 16
US-09-252-991A-28250
; Sequence 28250, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28250
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28250
Alignment Scores:
Pred. No.: 49.4 Length: 249
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 4 Gaps: 0
US-09-729-264-1 (1-1175) x US-09-252-991A-28250 (1-249)
QY 153 GCTTCAGCCTGGAGACGGTGC 130
DB 9 AlaSerProGlyArgArgCys 16
RESULT 17
US-09-489-039A-9994
; Sequence 9994, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9994
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9994
Alignment Scores:
Pred. No.: 49 Length: 271
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 4 Gaps: 0
US-09-729-264-1 (1-1175) x US-09-489-039A-9994 (1-271)
QY 629 CTTCGGGCTTCAGGCTCTCCA 606
DB 125 LeuAlaGlyLeuGlnAlaLeuPro 132
RESULT 18
US-09-252-991A-33090
; Sequence 33090, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 33090
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-33090
Alignment Scores:
Pred. No.: 46.8 Length: 460
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.06% Indels: 0
DB: 4 Gaps: 0
US-09-729-264-1 (1-1175) x US-09-252-991A-33090 (1-460)
QY 307 GGACATCAGATGCAGCTCCAGA 330
DB 244 GlyThrSerAspAlaAlaSerArg 251
RESULT 19
US-09-252-991A-24229
; Sequence 24229, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24229
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24229
Alignment Scores:
Pred. No.: 46.5 Length: 495
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.06% Indels: 0
DB: 4 Gaps: 0
US-09-729-264-1 (1-1175) x US-09-252-991A-24229 (1-495)
QY 1083 GCTGATCAGCTCCAGCCAGCCA 1106
DB 279 AlaAspGlnArgProProArgPro 286
RESULT 20
US-08-891-298-3
; Sequence 3, Application US/08891298
; Patent No. 6300488

```

```

; GENERAL INFORMATION:
; APPLICANT: Gage, Frederick H.
; APPLICANT: Sunr, Steven T.
; TITLE OF INVENTION: Modified Lepidopteran Receptors
; TITLE OF INVENTION: and Hybrid Multi-Functional Proteins for Use in Transcription
; TITLE OF INVENTION: and Transgene Expression Regulation
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,298
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-677-1409
; TELEFAX: 619-677-1465
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 606 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-891-298-3

Alignment Scores:
Pred. No.: 45.7 Length: 606
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-08-891-298-3 (1-606)
QY 853 CAGCAGCAGTTGCAGCCACAA 830
Db 135 GlnGlnLeuGlnProGlnGln 142

RESULT 21
US-08-653-648A-11
; Sequence 11, Application US/08653648A
; Patent No. 6379945
; GENERAL INFORMATION:
; APPLICANT: Jepson, Ian
; APPLICANT: Greenland, Andrew
; APPLICANT: Martinez, Alberto
; TITLE OF INVENTION: A Gene Switch
; FILE REFERENCE: PPD50047/US
; CURRENT APPLICATION NUMBER: US/08/653,648A
; CURRENT FILING DATE: 1996-05-24
; PRIOR APPLICATION NUMBER: GB 9510759.5
; PRIOR FILING DATE: 1995-05-26
; PRIOR APPLICATION NUMBER: GB 9605656.9

```

```

; PRIOR FILING DATE: 1996-03-18
; PRIOR APPLICATION NUMBER: GB 9513882.2
; PRIOR FILING DATE: 1995-07-07
; PRIOR APPLICATION NUMBER: GB 9517116.7
; PRIOR FILING DATE: 1995-08-24
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 606
; TYPE: PRT
; ORGANISM: Bombyx mori
; US-08-653-648A-11

Alignment Scores:
Pred. No.: 45.7 Length: 606
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-08-653-648A-11 (1-606)
QY 853 CAGCAGCAGTTGCAGCCACAA 830
Db 135 GlnGlnLeuGlnProGlnGln 142

RESULT 22
US-09-564-418-10
; Sequence 10, Application US/09564418
; Patent No. 6610828
; GENERAL INFORMATION:
; APPLICANT: Syngenta
; APPLICANT: Jepson, Ian
; APPLICANT: Martinez, Alberto
; APPLICANT: Greenland, Andrew James
; TITLE OF INVENTION: A GENE SWITCH
; FILE REFERENCE: 1392/4/3
; CURRENT APPLICATION NUMBER: US/09/564,418
; CURRENT FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: US 09/564,418
; PRIOR FILING DATE: 2000-05-03
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 606
; TYPE: PRT
; ORGANISM: Bombyx mori
; US-09-564-418-10

Alignment Scores:
Pred. No.: 45.7 Length: 606
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-564-418-10 (1-606)
QY 853 CAGCAGCAGTTGCAGCCACAA 830
Db 135 GlnGlnLeuGlnProGlnGln 142

RESULT 23
US-09-134-001C-4394
; Sequence 4394, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lyuu Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007

```

<p> APPLICANT: Shyjan, Andrew W.  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  CURRENT FILING DATE: 1998-08-13  PRIOR APPLICATION NUMBER: US 60/064,964  PRIOR FILING DATE: 1997-11-08  PRIOR APPLICATION NUMBER: US 60/055,779  PRIOR FILING DATE: 1997-08-14  NUMBER OF SEQ ID NOS: 5674  SEQ ID NO 4394  LENGTH: 629  TYPE: PRT  ORGANISM: Staphylococcus epidermidis  US-09-134-001C-4394 </p>	<p> Alignment Scores:  Pred. No.: 45.6  Score: 8.00  Percent Similarity: 100.00%  Best Local Similarity: 100.00%  Query Match: 2.11%  DB: 4 </p>	<p> Length: 629  Matches: 8  Conservative: 0  Mismatch: 0  Indels: 0  Gaps: 0 </p>
<p> US-09-729-264-1 (1-1175) x US-09-134-001C-4394 (1-629)  QY 1019 GGGAGGGAGAGAGCGGTCTGT 996  DB 444 GlyArgGluArgSerGlyValCys 451 </p>		
<p> RESULT 24  US-09-644-460-37  Sequence 37, Application US/09644460  Patent No. 6657053  GENERAL INFORMATION:  APPLICANT: Fisher, Paul B.  TITLE OF INVENTION: Reciprocal Subtraction Differential  FILE REFERENCE: 34587-C-PCT-USA  CURRENT FILING DATE: 2000-08-23  PRIOR APPLICATION NUMBER: PCT/US99/04323  PRIOR FILING DATE: 1999-02-26  PRIOR APPLICATION NUMBER: US 09/197,889  PRIOR FILING DATE: 1998-11-23  PRIOR APPLICATION NUMBER: US 09/185,115  PRIOR FILING DATE: 1998-11-03  PRIOR APPLICATION NUMBER: US 09/032,684  PRIOR FILING DATE: 1998-02-27  NUMBER OF SEQ ID NOS: 42  SOFTWARE: FastSeq for Windows Version 4.0  SEQ ID NO 37  LENGTH: 717  TYPE: PRT  ORGANISM: homo sapiens  US-09-644-460-37 </p>	<p> Alignment Scores:  Pred. No.: 45  Score: 8.00  Percent Similarity: 100.00%  Best Local Similarity: 100.00%  Query Match: 2.11%  DB: 4 </p>	<p> Length: 717  Matches: 8  Conservative: 0  Mismatch: 0  Indels: 0  Gaps: 0 </p>
<p> US-09-729-264-1 (1-1175) x US-09-644-460-37 (1-717)  QY 830 ACGAGCGGCGAGCAGCAGCG 807  DB 482 ThrThrAlaAlaAlaAlaAla 489 </p>		
<p> RESULT 25  US-09-035-648-18  Sequence 18, Application US/09035648  Patent No. 6100031 </p>		

;  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/001,951  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION:  
; APPLICATION NUMBER: 08/818,829  
; FILING DATE: 14-MAR-1997  
; APPLICATION NUMBER: 60/013,438  
; FILING DATE: 15-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/003001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-542-5070  
; TELEFAX: 617-542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 747 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-09-001-951-18  
  
Alignment Scores:  
Pred. No.: 44.9 Length: 747  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 3 Gaps: 0  
  
US-09-729-264-1 (1-1175) x US-09-001-951-18 (1-747)  
  
QY 1067 GCCACGCTACTGTTCTTTGTC 1044  
Db |||||||||||||||||||  
312 AlathrAlathrValSerLeuPhe 319  
  
RESULT 27  
US-08-818-829-18  
; Sequence 18, Application US/08818829  
; Patent No. 6458939  
; GENERAL INFORMATION:  
; APPLICANT: Shyjan, Andrew W.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL  
; TITLE OF INVENTION: GROWTH AND PROLIFERATION  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,829  
; FILING DATE: 14-MAR-1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/013,438  
; FILING DATE: 15-MAR-1996

;  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/003001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-542-5070  
; TELEFAX: 617-542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 747 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-818-829-18  
  
Alignment Scores:  
Pred. No.: 44.9 Length: 747  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 4 Gaps: 0  
  
US-09-729-264-1 (1-1175) x US-08-818-829-18 (1-747)  
  
QY 1067 GCCACGCTACTGTTCTTTGTC 1044  
Db |||||||||||||||||||  
312 AlathrAlathrValSerLeuPhe 319  
  
RESULT 28  
US-09-252-991A-27424  
; Sequence 27424, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27424  
; LENGTH: 751  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27424  
  
Alignment Scores:  
Pred. No.: 44.9 Length: 751  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 4 Gaps: 0  
  
US-09-729-264-1 (1-1175) x US-09-252-991A-27424 (1-751)  
  
QY 708 AGTTTACCGAGTTTAGTTTTC 731  
Db |||||||||||||||||||  
560 SerLeuProSerLeuGlyPheSer 567  
  
RESULT 29  
US-07-906-349A-6  
; Sequence 6, Application US/07906349A  
; Patent No. 5434064  
; GENERAL INFORMATION:  
; APPLICANT: Schlesinger, Joseph

```

Pred. No.: 44.6 Length: 808
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-134-000C-4643 (1-808)

QY 1079 CTGCTGAGGAGGCGCACAGTACT 1056
      |||||
Db 288 LeuValArgAlaThrAlaThr 295

RESULT 31
US-08-630-915A-37
; Sequence 37, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLKES, Dana M.
; APPLICANT: McCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-630-915A-37

Alignment Scores:
Pred. No.: 42.5 Length: 1400
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-08-630-915A-37 (1-1400)

```

## RESULT 32

US-09-060-854B-2  
; Sequence 2, Application US/09060854B  
; Patent No. 6642011  
; GENERAL INFORMATION:  
; APPLICANT: Estell, David Aaron  
; TITLE OF INVENTION: Human Protease and Use of Such Protease for Pharmaceutical  
; TITLE OF INVENTION: Applications and for Reducing the Allergenicity of No. 6642011-H  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: GC532  
; CURRENT APPLICATION NUMBER: US/09/060,854B  
; CURRENT FILING DATE: 1998-04-15  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1497  
; TYPE: PRT  
; ORGANISM: B. amyloliquefaciens  
US-09-060-854B-2

Alignment Scores:  
Pred. No.: 42.2 Length: 1497  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-060-854B-2 (1-1497)

QY 842 GCAGCCACAAACGCGCGCA 819  
Db 1342 AlaAlaThrThrThrAlaAla 1349  
|||||

## RESULT 33

US-09-627-650B-7  
; Sequence 7, Application US/09627650B  
; Patent No. 6406872  
; GENERAL INFORMATION:  
; APPLICANT: Bamber, Bruce  
; APPLICANT: Jorgensen, Erik  
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and  
; TITLE OF INVENTION: Methods Related Thereto  
; FILE REFERENCE: 21101.0009U3  
; CURRENT APPLICATION NUMBER: US/09/627,650B  
; CURRENT FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 09/436,063  
; PRIOR FILING DATE: 1999-11-08  
; PRIOR APPLICATION NUMBER: 60/107,727  
; PRIOR FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 2508  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-09-627-650B-7

Alignment Scores:  
Pred. No.: 40.4 Length: 2508  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-627-650B-7 (1-2508)

QY 830 ACCGCGCGCGCAGCAGCAGCG 807  
Db 2491 ThrThrAlaAlaAlaAlaAla 2498  
|||||

## RESULT 34

US-09-436-063C-7  
; Sequence 7, Application US/09436063C  
; Patent No. 6407210  
; GENERAL INFORMATION:  
; APPLICANT: Bamber, Bruce  
; APPLICANT: Jorgensen, Erik  
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and  
; TITLE OF INVENTION: Methods Related Thereto  
; FILE REFERENCE: P-1095corrected  
; CURRENT APPLICATION NUMBER: US/09/436,063C  
; CURRENT FILING DATE: 1999-11-08  
; PRIOR APPLICATION NUMBER: 60/107727  
; PRIOR FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 2508  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-09-436-063C-7

Alignment Scores:  
Pred. No.: 40.4 Length: 2508  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-436-063C-7 (1-2508)

QY 830 ACGACGCGCGCAGCAGCAGCG 807  
Db 2491 ThrThrAlaAlaAlaAlaAla 2498  
|||||

## RESULT 35

US-09-627-650B-3  
; Sequence 3, Application US/09627650B  
; Patent No. 6406872  
; GENERAL INFORMATION:  
; APPLICANT: Bamber, Bruce  
; APPLICANT: Jorgensen, Erik  
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and  
; TITLE OF INVENTION: Methods Related Thereto  
; FILE REFERENCE: 21101.0009U3  
; CURRENT APPLICATION NUMBER: US/09/627,650B  
; CURRENT FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 09/436,063  
; PRIOR FILING DATE: 1999-11-08  
; PRIOR APPLICATION NUMBER: 60/107,727  
; PRIOR FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 2544  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-09-627-650B-3

Alignment Scores:  
Pred. No.: 40.3 Length: 2544  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-627-650B-3 (1-2544)

QY 830 ACCGCGCGCGCAGCAGCAGCG 807  
Db 2527 ThrThrAlaAlaAlaAlaAla 2534  
|||||



Mon Sep 20 11:21:26 2004

us-09-729-264-1.oli.ra1

Db 2584 ThrThrAlaAlaAlaAlaAla 2591

```

RESULT 38
US-09-436-063C-9
; Sequence 9, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436,063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2601
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-9

Alignment Scores:
Pred. No.: 40.2 Length: 2601
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 4 Gaps: 0

```

US-09-729-264-1 (1-1175) x US-09-436-063C-9 (1-2601)

QY 830 ACGACGGCGGACGACGACG 807  
 Db 2584 ThrThrAlaAlaAlaAlaAla 2591

```

RESULT 39
US-08-425-069-56
; Sequence 56, Application US/08425069
; Patent No. 5728810
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Hinman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5728810th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/425,069
; FILING DATE: 19-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-106P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050

```

```

RESULT 36
US-09-436-063C-3
; Sequence 3, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436,063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-3

Alignment Scores:
Pred. No.: 40.3 Length: 2544
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 4 Gaps: 0

```

US-09-729-264-1 (1-1175) x US-09-436-063C-3 (1-2544)

QY 830 ACGACGGCGGACGACGACG 807  
 Db 2527 ThrThrAlaAlaAlaAlaAla 2534

```

RESULT 37
US-09-627-650B-9
; Sequence 9, Application US/09627650B
; Patent No. 6406872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: 21101.000903
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2601
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-9

Alignment Scores:
Pred. No.: 40.2 Length: 2601
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 4 Gaps: 0

```

US-09-729-264-1 (1-1175) x US-09-627-650B-9 (1-2601)

QY 830 ACGACGGCGGACGACGACG 807

```
/
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 56:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 27 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: not relevant
/ TOPOLOGY: not relevant
/ MOLECULE TYPE: peptide
/ HYPOTHETICAL: NO
/ FRAGMENT TYPE: internal
/ ORIGINAL SOURCE:
/ ORGANISM: nephila clavipes
/ FEATURE:
/ NAME/KEY: Peptide
/ LOCATION: 1..27
/ OTHER INFORMATION: /label= silk1_repeat
US-08-425-069-56

Alignment Scores:
Pred. No.: 519 Length: 27
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x US-08-425-069-56 (1-27)
QY 808 GCTGCTGCTGCTGCGCGTC 828
Db 10 AlaAlaAlaAlaAlaVal 16

RESULT 40
US-08-317-844B-56
/ Sequence 56, Application US/08317844B
/ Patent No. 5989894
/ GENERAL INFORMATION:
/ APPLICANT: Lewis, Randolph V.
/ APPLICANT: Xu, Ming
/ APPLICANT: Hinman, Michael B.
/ TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
/ TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
/ TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
/ NUMBER OF SEQUENCES: 62
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Birch, Stewart, Kolasch & Birch
/ STREET: 301 No. 5989894th Washington Street
/ CITY: Falls Church
/ STATE: Virginia
/ COUNTRY: U.S.A.
/ ZIP: 22046
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/317,844B
/ FILING DATE: 04-OCT-1994
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Murphy Jr., Gerald M
/ REGISTRATION NUMBER: 28,977
/ REFERENCE/DOCKET NUMBER: 1447-105P
/ TELEPHONE: (703) 241-1300
/ TELEFAX: (703) 241-2848
/ TELEX: 248345
/ INFORMATION FOR SEQ ID NO: 56:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 27 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: not relevant
```

```
/
/ TOPOLOGY: not relevant
/ MOLECULE TYPE: peptide
/ HYPOTHETICAL: NO
/ FRAGMENT TYPE: internal
/ ORIGINAL SOURCE:
/ ORGANISM: nephila clavipes
/ FEATURE:
/ NAME/KEY: Peptide
/ LOCATION: 1..27
/ OTHER INFORMATION: /label= silk1_repeat
US-08-317-844B-56

Alignment Scores:
Pred. No.: 519 Length: 27
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x US-08-317-844B-56 (1-27)
QY 808 GCTGCTGCTGCTGCGCGTC 828
Db 10 AlaAlaAlaAlaAlaVal 16

RESULT 41
US-08-905-223-390
/ Sequence 390, Application US/08905223
/ Patent No. 6222029
/ GENERAL INFORMATION:
/ APPLICANT: Edwards, Jean-Baptiste D.
/ APPLICANT: Duclert, Aymeric
/ APPLICANT: Lacroix, Bruno
/ TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS
/ NUMBER OF SEQUENCES: 503
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Knobbe, Martens, Olson & Bear
/ STREET: 501 West Broadway
/ CITY: San Diego
/ STATE: California
/ COUNTRY: USA
/ ZIP: 92101-3505
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy Disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: Win95
/ SOFTWARE: Word
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/905,223
/ FILING DATE:
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Israelsen, Ned A.
/ REGISTRATION NUMBER: 29,655
/ REFERENCE/DOCKET NUMBER:
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (619) 235-8550
/ TELEFAX: (619) 235-0176
/ INFORMATION FOR SEQ ID NO: 390:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 46 amino acids
/ TYPE: AMINO ACID
/ TOPOLOGY: LINEAR
/ MOLECULE TYPE: PROTEIN
/ ORIGINAL SOURCE:
/ ORGANISM: Homo Sapiens
/ TISSUE TYPE: Brain
/ FEATURE:
/ NAME/KEY: sig peptide
/ LOCATION: -34..-1
/ IDENTIFICATION METHOD: Von Heijne matrix
/ OTHER INFORMATION: score 3.6
```

```
OTHER INFORMATION: seq IKSSWISSLASG/IP
US-08-905-223-390

Alignment Scores:
Pred. No.: 496 Length: 46
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 3 Gaps: 0

US-09-729-264-1 (1-1175) x US-08-905-223-390 (1-46)

QY 717 AGTTTAGTTTTCATTCGCT 737
|||||
Db 38 SerLeuGlyPheSerLeuPro 44

RESULT 42
US-09-117-121-30
; Sequence 30, Application US/09117121
; Patent No. 6307020
; GENERAL INFORMATION:
; APPLICANT: Hew, Choy
; APPLICANT: Gong, Zhiyuan
; TITLE OF INVENTION: Intracellular Antifreeze Polypeptides
; TITLE OF INVENTION: and Nucleic Acids
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/117,121
; FILING DATE: 20-NOV-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CA97/00062
; FILING DATE: 30-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 016252-001610US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-117-121-30

Alignment Scores:
Pred. No.: 489 Length: 54
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.84% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-117-121-30 (1-54)

QY 827 ACGGCGCAGCAGCAGCG 807
|||||

Db 22 ThrAlaAlaAlaAlaAla 28

RESULT 43
US-09-328-352-7668
; Sequence 7668, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7668
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7668

Alignment Scores:
Pred. No.: 483 Length: 62
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.84% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-328-352-7668 (1-62)

QY 43 AAAATCACAGCCCATCTCT 23
|||||
Db 23 LysIleThrSerProIleSer 29

RESULT 44
US-09-238-303-15
; Sequence 15, Application US/09238303B
; Patent No. 6284253
; GENERAL INFORMATION:
; APPLICANT: Barr, Margaret C.
; TITLE OF INVENTION: No. 6284253el Feline Immunodeficiency Virus Nucleotide Sequence
; FILE REFERENCE: 18617.0059
; CURRENT APPLICATION NUMBER: US/09/238,303B
; CURRENT FILING DATE: 1999-01-28
; EARLIER FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 15
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: protein encoded by the open reading frame (orfe) of a
; OTHER INFORMATION: recombinant viral clone constructed from the genomic DNA of a
; OTHER INFORMATION: Pallas's cat feline immunodeficiency virus
US-09-238-303-15

Alignment Scores:
Pred. No.: 481 Length: 65
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 3 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-238-303-15 (1-65)

QY 873 AGAAGAAAAGAGGATTCGT 893
|||||
Db 24 ArgArgLysArgGlyPheArg 30

RESULT 45
US-09-946-239-15
```

; Sequence 15, Application US/09946239  
; Patent No. 6579527  
; GENERAL INFORMATION:  
; APPLICANT: Barr, Margaret C.  
; TITLE OF INVENTION: No. 6579527el Feline Immunodeficiency Virus Nucleotide and  
; TITLE OF INVENTION: Polypeptide Sequences  
; FILE REFERENCE: 18617.0059  
; CURRENT APPLICATION NUMBER: US/09/946,239  
; CURRENT FILING DATE: 2001-09-04  
; PRIOR APPLICATION NUMBER: US 09/238,303, US 60/072,927  
; PRIOR FILING DATE: 1999-01-28, 1998-01-29  
; NUMBER OF SEQ ID NOS: 17  
; SEQ ID NO 15  
; LENGTH: 65  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: protein encoded by the open reading frame (orfe) of a  
; OTHER INFORMATION: recombinant viral clone constructed from the genomic DNA of a  
; OTHER INFORMATION: Pallas's cat feline immunodeficiency virus  
US-09-946-239-15

Alignment Scores:  
Pred. No.: 481 Length: 65  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.80% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-946-239-15 (1-65)

Qy 873 AGAGAGAAACAGGATTTCGT 893  
Db 24 ArgArgLysArgGlyPheArg 30

Search completed: September 18, 2004, 23:04:51  
Job time : 37.9373 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 18, 2004, 22:53:37 ; Search time 63.128 Seconds  
(without alignments)  
11954.499 Million cell updates

Title: US-09-729-264-1

Perfect score: 389

Perfect score: 100  
Sequence: 1 ctgtctgcccatctgaataa.....gtaataactgtagtatag 1175

Scoring table:

Scoring table:			
	60.0	Xgapext	60.0
Xgapop	60.0	Ygapext	60.0
Ygapop	6.0	Fgapext	7.0
Fgapop	6.0	Delext	7.0

Searched: 1342398 seqs, 321133274 residues

word size:

Total number of hits satisfying chosen parameters: 2128678

Minimum DB seq length: 25

Maximum DB seq length: 2000000000  
Maximum DB seq length: 2000000000

Post-processing: Listing first 135 summaries

Command line parameters:

```

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q/cg2n2_1/USTPO.spool_p/US09729264/runat_17092004_155108_2424/app_query.fasta_1.4117
-Q/cg2n2_1/USTPO.spool_AA_QFTWT=fastan_SUFFIX=oli.rapb_MINMATCH=0.1
-DB=Published Applications
-LOOPC1=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=135 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=0
-ALIGN=45 -MODE=LOCAL -OUTFWT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=35
-MAXLEN=200000000 -USER=US09729264 @CGN 1 1 323 @runat_17092004_155108_2424
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -WAIT TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-E2A POP=6 -EG3EXT=7 -YGAPOP=60 -YGAPEXT=60 -DELEXT=7

```

Database : Published Applications AA:\*\*

Database : Published Applications -AA:

1:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp:
2:	/cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp:
3:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp:
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp:
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp:
6:	/cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp:
7:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp:
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp:
9:	/cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp:
10:	/cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp:
11:	/cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp:
12:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp:
13:	/cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp:
14:	/cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp:
15:	/cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp:
16:	/cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp:
17:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp:
18:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	DB Length	ID	Description
1	100	100	100	100	100
2	95	95	95	95	95
3	90	90	90	90	90
4	85	85	85	85	85
5	80	80	80	80	80
6	75	75	75	75	75
7	70	70	70	70	70
8	65	65	65	65	65
9	60	60	60	60	60
10	55	55	55	55	55
11	50	50	50	50	50
12	45	45	45	45	45
13	40	40	40	40	40
14	35	35	35	35	35
15	30	30	30	30	30
16	25	25	25	25	25
17	20	20	20	20	20
18	15	15	15	15	15
19	10	10	10	10	10
20	5	5	5	5	5





```
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 561
; LENGTH: 1060
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-561
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 561
; LENGTH: 1060
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-561
Alignment Scores:
Pred. No.: 0.0547 Length: 1060
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0
US-09-729-264-1 (1-1175) x US-10-184-634-561 (1-1060)
QY 842 GCAGCCACAACAGCAGCGGCGCAGCAGCAGCG 807
Db 254 AlaAlaThrThrThrAlaAlaAlaAlaAla 265
RESULT 5
US-10-063-685-165
; Sequence 165, Application US/10063685
; Publication No. US20030180909A1
; GENERAL INFORMATION:
; APPLICANT: Eaton,Dan L.
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,Christopher J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,685
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 165
; LENGTH: 1060
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-063-685-165
Alignment Scores:
Pred. No.: 0.0547 Length: 1060
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0
US-09-729-264-1 (1-1175) x US-10-063-685-165 (1-1060)
```

```
QY 842 GCAGCCACAACAGCAGCGGCGCAGCAGCAGCG 807
Db 254 AlaAlaThrThrThrAlaAlaAlaAlaAla 265
RESULT 6
US-10-184-644-591
; Sequence 591, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Revin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 591
; LENGTH: 2037
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-591
Alignment Scores:
Pred. No.: 0.0507 Length: 2037
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0
US-09-729-264-1 (1-1175) x US-10-184-644-591 (1-2037)
QY 842 GCAGCCACAACAGCAGCGGCGCAGCAGCAGCG 807
Db 193 AlaAlaThrThrThrAlaAlaAlaAlaAla 204
RESULT 7
US-10-184-634-591
; Sequence 591, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 591
; LENGTH: 2037
```



```
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-591

Alignment Scores:
Pred. No.: 0.0507 Length: 2037
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-184-634-591 (1-2037)

QY 842 GCAGCCACAAACGCGCGGCGAGCAGCAGCG 807
Db 193 AlaAlaThrThrThrAlaAlaAlaAlaAla 204

RESULT 8
US-10-142-426-137
; Sequence 137, Application US/10142426
; Publication No. US20040048333A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 137
; LENGTH: 2207
; TYPE: DNA
; ORGANISM: Homo Sapien
; NAME/KEY: unsure
; LOCATION: 2153, 2160
; OTHER INFORMATION: unknown base
US-10-123-155-137

Alignment Scores:
Pred. No.: 0.0503 Length: 2207
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-123-155-137 (1-2207)

QY 842 GCAGCCACAAACGCGCGGCGAGCAGCAGCG 807
Db 2161 AlaAlaThrThrThrAlaAlaAlaAlaAla 2172

RESULT 10
US-10-146-731-137
; Sequence 137, Application US/10146731
; Publication No. US20030129692A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
```

```
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C323
; CURRENT APPLICATION NUMBER: US/10/146,731
; CURRENT FILING DATE: 2002-05-15
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 137
; LENGTH: 2207
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2153, 2160
; OTHER INFORMATION: unknown base
US-10-146-731-137

Alignment Scores:
Pred. No.: 0.0503 Length: 2207
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-146-731-137 (1-2207)

QY 842 GCAGCCACACACAGCGCGGCGAGCAGCAGCG 807
Db 2161 AlaAlaThrThrThrAlaAlaAlaAlaAla 2172

RESULT 11
US-10-140-472-137
; Sequence 137, Application US/10140472
; Publication No. US20030138888A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C168
; CURRENT APPLICATION NUMBER: US/10/140,472
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 137
; LENGTH: 2207
; TYPE: DNA
; ORGANISM: Homo Sapien
; NAME/KEY: unsure
; FEATURE:
; LOCATION: 2153, 2160
; OTHER INFORMATION: unknown base
US-10-140-472-137

Alignment Scores:
Pred. No.: 0.0503 Length: 2207
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-140-472-137 (1-2207)

QY 842 GCAGCCACACACAGCGCGGCGAGCAGCAGCG 807
Db 2161 AlaAlaThrThrThrAlaAlaAlaAlaAla 2172

RESULT 12
US-10-141-761-137
; Sequence 137, Application US/10141761
; Publication No. US20030148432A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 137
; LENGTH: 2207
; TYPE: DNA
; ORGANISM: Homo Sapien
; NAME/KEY: unsure
; FEATURE:
; LOCATION: 2153, 2160
; OTHER INFORMATION: unknown base
US-10-141-761-137

Alignment Scores:
Pred. No.: 0.0503 Length: 2207
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-141-761-137 (1-2207)

QY 842 GCAGCCACACACAGCGCGGCGAGCAGCAGCG 807
Db 2161 AlaAlaThrThrThrAlaAlaAlaAlaAla 2172

RESULT 13
US-10-142-885-137
; Sequence 137, Application US/10142885
; Publication No. US20030157604A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
```

```
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-140-472-137 (1-2207)

QY 842 GCAGCCACACACAGCGCGGCGAGCAGCAGCG 807
Db 2161 AlaAlaThrThrThrAlaAlaAlaAlaAla 2172

RESULT 12
US-10-141-761-137
; Sequence 137, Application US/10141761
; Publication No. US20030148432A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 137
; LENGTH: 2207
; TYPE: DNA
; ORGANISM: Homo Sapien
; NAME/KEY: unsure
; FEATURE:
; LOCATION: 2153, 2160
; OTHER INFORMATION: unknown base
US-10-141-761-137

Alignment Scores:
Pred. No.: 0.0503 Length: 2207
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-141-761-137 (1-2207)

QY 842 GCAGCCACACACAGCGCGGCGAGCAGCAGCG 807
Db 2161 AlaAlaThrThrThrAlaAlaAlaAlaAla 2172

RESULT 13
US-10-142-885-137
; Sequence 137, Application US/10142885
; Publication No. US20030157604A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
```



Qy 842 GCAGCCACAACAGCGGGCGGAGCAGCAGCG 807  
Db 2161 AlaAlaThrThrThrAlaAlaAlaAlaAla 2172

## RESULT 16

US-10-140-923-137

; Sequence 137, Application US/10140923

; Publication No. US20030207355A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tamas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3330R1C188

; CURRENT APPLICATION NUMBER: US/10/140,923

; CURRENT FILING DATE: 2002-05-07

; Prior Application removed - See Palm or File Wrapper

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 137

; LENGTH: 2207

; TYPE: DNA

; ORGANISM: Homo Sapien

; FEATURE:

; NAME/KEY: unsure

; LOCATION: 2153, 2160

; OTHER INFORMATION: unknown base

US-10-140-923-137

Alignment Scores:

Pred. No.:	0.0503	Length:	2207
Score:	12.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.16%	Indels:	0
DB:	15	Gaps:	0

US-09-729-264-1 (1-1175) x US-10-140-923-137 (1-2207)

Qy 842 GCAGCCACAACAGCGGGCGGAGCAGCAGCG 807  
Db 2161 AlaAlaThrThrThrAlaAlaAlaAlaAla 2172

## RESULT 17

US-10-141-756-137

; Sequence 137, Application US/10141756

; Publication No. US20030207355A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tamas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C200  
; CURRENT APPLICATION NUMBER: US/10/141,756  
; CURRENT FILING DATE: 2002-05-08  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 137  
; LENGTH: 2207  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 2153, 2160  
; OTHER INFORMATION: unknown base  
US-10-141-756-137

Alignment Scores:

Pred. No.:	0.0503	Length:	2207
Score:	12.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.16%	Indels:	0
DB:	15	Gaps:	0

US-09-729-264-1 (1-1175) x US-10-141-756-137 (1-2207)

Qy 842 GCAGCCACAACAGCGGGCGGAGCAGCAGCG 807  
Db 2161 AlaAlaThrThrThrAlaAlaAlaAlaAla 2172

## RESULT 18

US-10-141-759-137

; Sequence 137, Application US/10141759

; Publication No. US20030207361A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tamas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3330R1C197

; CURRENT APPLICATION NUMBER: US/10/141,759

; CURRENT FILING DATE: 2002-05-08

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 137

; LENGTH: 2207

; TYPE: DNA

; ORGANISM: Homo Sapien

; FEATURE:

; NAME/KEY: unsure

; LOCATION: 2153, 2160

```
; OTHER INFORMATION: unknown base
US-10-141-759-137

Alignment Scores:
Pred. No.: 0.0503 Length: 2207
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 15 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-141-759-137 (1-2207)
QY 842 GCAGCCACAAACAGCGCGGCAGCAGCAGCG 807
Db 2161 AlaAlaThrThrThrAlaAlaAlaAla 2172

RESULT 19
US-10-140-805-137
; Sequence 137, Application US/10140805
; Publication No. US20030207419A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C176
; CURRENT APPLICATION NUMBER: US/10/140,805
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 137
; LENGTH: 2207
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2153, 2160
; OTHER INFORMATION: unknown base
US-10-140-864-137

Alignment Scores:
Pred. No.: 0.0503 Length: 2207
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 15 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-140-864-137 (1-2207)
QY 842 GCAGCCACAAACAGCGCGGCAGCAGCAGCG 807
Db 2161 AlaAlaThrThrThrAlaAlaAlaAla 2172

RESULT 21
US-10-142-426-331
; Sequence 331, Application US/10142426
; Publication No. US20040048333A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
```

```
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C224
; CURRENT APPLICATION NUMBER: US/10/142,426
; CURRENT FILING DATE: 2002-05-09
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 331
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-142-426-331

Alignment Scores:
Pred. No.: 0.0496 Length: 2477
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 12 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-142-426-331 (1-2477)
QY 842 GCAGCCACAAACGAGCGCGGAGCAGCAGCG 807
Db 518 AlaAlaThrThrThrAlaAlaAlaAlaAla 529

RESULT 22
US-10-123-155-331
; Sequence 331, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C30
; CURRENT APPLICATION NUMBER: US/10/146,731
; CURRENT FILING DATE: 2002-05-15
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 331
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-146-731-331

Alignment Scores:
Pred. No.: 0.0496 Length: 2477
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-146-731-331 (1-2477)
QY 842 GCAGCCACAAACGAGCGCGGAGCAGCAGCG 807
Db 518 AlaAlaThrThrThrAlaAlaAlaAlaAla 529

RESULT 24
US-10-140-472-331
; Sequence 331, Application US/10140472
; Publication No. US20030138888A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
```

```
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P333OR1C168
; CURRENT APPLICATION NUMBER: US/10/140,472
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 331
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-472-331
Alignment Scores:
Pred. No.: 0.0496 Length: 2477
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0
US-09-729-264-1 (1-1175) x US-10-140-472-331 (1-2477)
Qy 842 GCAGCCACAAACGACGCGCGCAGCAGCAGCG 807
Db 518 AlaAlaThrThrThrAlaAlaAlaAlaAla 529
RESULT 25
US-10-141-761-331
; Sequence 331, Application US/10141761
; Publication No. US20030149432A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P333OR1C248
; CURRENT APPLICATION NUMBER: US/10/142,885
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 331
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-142-885-331
Alignment Scores:
Pred. No.: 0.0496 Length: 2477
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0
US-09-729-264-1 (1-1175) x US-10-142-885-331 (1-2477)
Qy 842 GCAGCCACAAACGACGCGCGCAGCAGCAGCG 807
Db 518 AlaAlaThrThrThrAlaAlaAlaAlaAla 529
RESULT 26
US-10-142-885-331
; Sequence 331, Application US/10142885
; Publication No. US20030157604A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P333OR1C248
; CURRENT APPLICATION NUMBER: US/10/142,885
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 331
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-142-885-331
Alignment Scores:
Pred. No.: 0.0496 Length: 2477
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0
US-09-729-264-1 (1-1175) x US-10-142-885-331 (1-2477)
Qy 842 GCAGCCACAAACGACGCGCGCAGCAGCAGCG 807
Db 518 AlaAlaThrThrThrAlaAlaAlaAlaAla 529
RESULT 27
US-10-138-790-331
; Sequence 331, Application US/10158790
; Publication No. US20030180879A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
```

```
Db 518 AlaAlaThrThrThrAlaAlaAlaAlaAla 529
RESULT 26
US-10-142-885-331
; Sequence 331, Application US/10142885
; Publication No. US20030157604A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P333OR1C248
; CURRENT APPLICATION NUMBER: US/10/142,885
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 331
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-142-885-331
Alignment Scores:
Pred. No.: 0.0496 Length: 2477
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0
US-09-729-264-1 (1-1175) x US-10-142-885-331 (1-2477)
Qy 842 GCAGCCACAAACGACGCGCGCAGCAGCAGCG 807
Db 518 AlaAlaThrThrThrAlaAlaAlaAlaAla 529
RESULT 27
US-10-138-790-331
; Sequence 331, Application US/10158790
; Publication No. US20030180879A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
```

```
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C448
; CURRENT APPLICATION NUMBER: US/10/158,790
; CURRENT FILING DATE: 2002-05-30
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 331
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-158-790-331

Alignment Scores:
Pred. No.: 0.0496 Length: 2477
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-158-790-331 (1-2477)
QY 842 GCAGCCACACAGCGCGGCGAGCAGCAGCG 807
Db 518 AlaAlaThrThrThrAlaAlaAlaAlaAla 529

RESULT 28
US-10-137-871-331
; Sequence 331, Application US/10137871
; Publication No. US20030207350A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin X
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C153
; CURRENT APPLICATION NUMBER: US/10/137,871
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 331
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-137-871-331

Alignment Scores:
Pred. No.: 0.0496 Length: 2477
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 15 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-137-871-331 (1-2477)
```

```
QY 842 GCAGCCACACAGCGCGGCGAGCAGCAGCG 807
Db 518 AlaAlaThrThrThrAlaAlaAlaAlaAla 529

RESULT 29
US-10-140-923-331
; Sequence 331, Application US/10140923
; Publication No. US20030207355A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C188
; CURRENT APPLICATION NUMBER: US/10/140,923
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 331
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-923-331

Alignment Scores:
Pred. No.: 0.0496 Length: 2477
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 15 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-140-923-331 (1-2477)
QY 842 GCAGCCACACAGCGCGGCGAGCAGCAGCG 807
Db 518 AlaAlaThrThrThrAlaAlaAlaAlaAla 529

RESULT 30
US-10-141-756-331
; Sequence 331, Application US/10141756
; Publication No. US20030207359A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
```



```

; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C200
; CURRENT APPLICATION NUMBER: US/10/141,756
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 331
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-141-756-331

Alignment Scores:
Pred. No.: 0.0496 Length: 2477
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 15 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-141-756-331 (1-2477)

QY 842 GCAGCCACAACAGCGCGGCGAGCAGCAGCG 807
Db 518 AlaaAlaThrThrThrThrAlaAlaAlaAlaAla 529

RESULT 31
US-10-141-759-331
; Sequence 331, Application US/10141759
; Publication No. US20030207361A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C197
; CURRENT APPLICATION NUMBER: US/10/141,759
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 331
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-141-759-331

Alignment Scores:
Pred. No.: 0.0496 Length: 2477
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 15 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-141-759-331 (1-2477)

```

```
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C184
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 331
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-864-331

Alignment Scores:
Pred. No.: 0.0496 Length: 2477
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.16% Indels: 0
DB: 15 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-140-864-331 (1-2477)
QY 842 GCAGCCACAACAGCGGCGGAGCAGCAGCG 807
Db 518 AlaAlaThrThrThrThrAlaAlaAlaAlaAla 529

RESULT 34
US-09-791-279-105
; Sequence 105, Application US/09791279
; Publication No. US20030050456A1
; GENERAL INFORMATION:
; APPLICANT: Vogel, Gabriel
; APPLICANT: Wood, Linda S.
; APPLICANT: Parodi, Luis
; APPLICANT: Lind, Peter
; TITLE OF INVENTION: No. US20030050456A1 G Protein-Coupled Receptors
; FILE REFERENCE: 00048 US1
; CURRENT APPLICATION NUMBER: US/09/791,279
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/184,715
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184725
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,712
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,606
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,602
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,604
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,822
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,710
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,689
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,690
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/184,716
; PRIOR FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 105
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-279-105
```

```
Alignment Scores:
Pred. No.: 0.508 Length: 524
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.89% Indels: 0
DB: 10 Gaps: 0

US-09-729-264-1 (1-1175) x US-09-791-279-105 (1-524)
QY 839 GCCACAACAGCGGCGGAGCAGCAGCG 807
Db 432 AlaThrThrThrThrAlaAlaAlaAlaAla 442

RESULT 35
US-10-063-685-17
; Sequence 17, Application US/10063685
; Publication No. US20030180909A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,685
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 17
; LENGTH: 1672
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-063-685-17

Alignment Scores:
Pred. No.: 0.444 Length: 1672
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.89% Indels: 0
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-063-685-17 (1-1672)
QY 842 GCAGCCACAACAGCGGCGGAGCAGCAGCA 810
Db 1272 AlaAlaThrThrThrThrAlaAlaAlaAla 1282

RESULT 36
US-10-142-426-189
; Sequence 189, Application US/10142426
; Publication No. US20040048333A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
```

; APPLICANT: Tumas,Daniel  
; APPLICANT: Watanabe,Colin K  
; APPLICANT: Wood,William  
; APPLICANT: Zhang,Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C224

; CURRENT APPLICATION NUMBER: US/10/142,426

; CURRENT FILING DATE: 2002-05-09

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 189

; LENGTH: 2150

; TYPE: DNA

; ORGANISM: Homo Sapien

US-10-142-426-189

Alignment Scores:

Pred. No.:	0.432	Length:	2150
Score:	11.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.89%	Indels:	0
DB:	12	Gaps:	0

US-09-729-264-1 (1-1175) x US-10-142-426-189 (1-2150)

QY 839 GCCACAAACAGCGCGGCAGCAGCAGCG 807

Db 2089 AlathrThrThrAlaAlaAlaAla 2099

RESULT 37

US-10-123-155-189

; Sequence 189, Application US/10123155

; Publication No. US20030068794A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3330R1C30

; CURRENT APPLICATION NUMBER: US/10/123,155

; CURRENT FILING DATE: 2002-04-15

; Prior Application removed - See Palm or File Wrapper

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 189

; LENGTH: 2150

; TYPE: DNA

; ORGANISM: Homo Sapien

US-10-123-155-189

Alignment Scores:

Pred. No.:	0.432	Length:	2150
Score:	11.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.89%	Indels:	0
DB:	14	Gaps:	0

US-09-729-264-1 (1-1175) x US-10-123-155-189 (1-2150)

QY 839 GCCACAAACAGCGCGGCAGCAGCAGCG 807

Db 2089 AlathrThrThrAlaAlaAlaAlaAla 2099

RESULT 38

US-10-146-731-189

; Sequence 189, Application US/10146731

; Publication No. US20030129692A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3330R1C323

; CURRENT APPLICATION NUMBER: US/10/146,731

; CURRENT FILING DATE: 2002-05-15

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 189

; LENGTH: 2150

; TYPE: DNA

; ORGANISM: Homo Sapien

US-10-146-731-189

Alignment Scores:

Pred. No.:	0.432	Length:	2150
Score:	11.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.89%	Indels:	0
DB:	14	Gaps:	0

US-09-729-264-1 (1-1175) x US-10-146-731-189 (1-2150)

QY 839 GCCACAAACAGCGCGGCAGCAGCAGCG 807

Db 2089 AlathrThrThrAlaAlaAlaAlaAla 2099

RESULT 39

US-10-140-472-189

; Sequence 189, Application US/10140472

; Publication No. US2003013888A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

```
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C168
; CURRENT APPLICATION NUMBER: US/10/140,472
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 189
; LENGTH: 2150
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-472-189

Alignment Scores:
Pred. No.: 0.432 Length: 2150
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.89% Indels: 0
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-140-472-189 (1-2150)
QY 839 GCCACACACGCGCGCGCAGCAGCAGCG 807
Db 2089 AlaThrThrThrAlaAlaAlaAlaAla 2099

RESULT 40
US-10-141-761-189
; Sequence 189, Application US/10141761
; Publication No. US20030148432A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C248
; CURRENT APPLICATION NUMBER: US/10/142,885
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 189
; LENGTH: 2150
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-142-885-189

Alignment Scores:
Pred. No.: 0.432 Length: 2150
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.89% Indels: 0
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-142-885-189 (1-2150)
QY 839 GCCACACACGCGCGCGCAGCAGCAGCG 807
Db 2089 AlaThrThrThrAlaAlaAlaAlaAla 2099

RESULT 42
US-10-158-790-189
; Sequence 189, Application US/10158790
; Publication No. US20030180879A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
```

```
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-141-761-189 (1-2150)
QY 839 GCCACACACGCGCGCGCAGCAGCAGCG 807
Db 2089 AlaThrThrThrAlaAlaAlaAlaAla 2099

RESULT 41
US-10-142-885-189
; Sequence 189, Application US/10142885
; Publication No. US20030157604A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C248
; CURRENT APPLICATION NUMBER: US/10/142,885
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 189
; LENGTH: 2150
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-142-885-189

Alignment Scores:
Pred. No.: 0.432 Length: 2150
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.89% Indels: 0
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-142-885-189 (1-2150)
QY 839 GCCACACACGCGCGCGCAGCAGCAGCG 807
Db 2089 AlaThrThrThrAlaAlaAlaAlaAla 2099

RESULT 42
US-10-158-790-189
; Sequence 189, Application US/10158790
; Publication No. US20030180879A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
```

; APPLICANT: Smith,Victoria  
; APPLICANT: Stewart,Timothy A.  
; APPLICANT: Tumas,Daniel  
; APPLICANT: Watanabe,Colin K  
; APPLICANT: Wood,William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C448  
; CURRENT APPLICATION NUMBER: US/10/158,790  
; CURRENT FILING DATE: 2002-05-30  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 189  
; LENGTH: 2150  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-158-790-189

Alignment Scores:  
Pred. No.: 0.432 Length: 2150  
Score: 11.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.89% Indels: 0  
DB: 14 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-158-790-189 (1-2150)

Qy 839 GCCACAACAGCGCGGCGACGACGACGCG 807  
Db 2089 AlathrThrThrAlaAlaAlaAlaAla 2099

## RESULT 43

US-10-137-871-189  
; Sequence 189, Application US/10137871  
; Publication No. US20030207350A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3330R1C153  
; CURRENT APPLICATION NUMBER: US/10/137,871

; CURRENT FILING DATE: 2002-05-03  
; Prior Application removed - See Palm or File Wrapper

; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 189

; LENGTH: 2150

; TYPE: DNA

; ORGANISM: Homo Sapien  
US-10-137-871-189

Alignment Scores:  
Pred. No.: 0.432 Length: 2150  
Score: 11.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.89% Indels: 0  
DB: 15 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-137-871-189 (1-2150)

Qy 839 GCCACAACAGCGCGGCGACGACGACGCG 807  
Db 2089 AlathrThrThrAlaAlaAlaAlaAla 2099

## RESULT 44

US-10-140-923-189  
; Sequence 189, Application US/10140923  
; Publication No. US20030207355A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3330R1C188  
; CURRENT APPLICATION NUMBER: US/10/140,923

; CURRENT FILING DATE: 2002-05-07  
; Prior Application removed - See Palm or File Wrapper

; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 189

; LENGTH: 2150

; TYPE: DNA

; ORGANISM: Homo Sapien  
US-10-140-923-189

## Alignment Scores:

Pred. No.: 0.432 Length: 2150  
Score: 11.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.89% Indels: 0  
DB: 15 Gaps: 0

US-09-729-264-1 (1-1175) x US-10-140-923-189 (1-2150)

Qy 839 GCCACAACAGCGCGGCGACGACGACGCG 807  
Db 2089 AlathrThrThrAlaAlaAlaAlaAla 2099

## RESULT 45

US-10-141-756-189

; Sequence 189, Application US/10141756  
; Publication No. US20030207359A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.

```

; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tamas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C200
; CURRENT APPLICATION NUMBER: US/10/141,756
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 189
; LENGTH: 2150
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-141-756-189

```

```

Alignment Scores:
Pred. No.: 0.432 Length: 2150
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.89% Indels: 0
DB: 15 Gaps: 0

```

US-09-729-264-1 (1-1175) x US-10-141-756-189 (1-2150)

```

QY 839 GCCACACACGACGCGCGCAGCAGCAGCG 807
Db 2089 AlaThrThrThrAlaAlaAlaAlaAla 2099

```

Search completed: September 18, 2004, 23:12:02  
Job time : 122.128 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 18, 2004, 22:45:42 ; Search time 20.1682 Seconds  
(without alignments)  
11208.268 Million cell updates/sec

Title: US-09-729-264-1

Perfect score: 389

Sequence: 1 ctgtctgccatctgaataa.....gtaatacaactgtagtatag 1175

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Word size: 0

Total number of hits satisfying chosen parameters: 557252

Minimum DB seq length: 25

Maximum DB seq length: 200000000

Post-processing: Listing first 135 summaries

Command line parameters: -DEV=xlp

-Q=/cgn2.1/USPTO\_spool\_p/US09729264/runat.17092004.155107.2371/app\_query.fasta.1.4117  
-DB=PIR78 -QFMT=fastan -SUFFIX=oli.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.gdi -LIST=135  
-DOCALIGN=200 -THR SCORE=quality -THR MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=plo  
-NORM=ext -HEAPSIZE=500 -MINLEN=25 -MAXLEN=200000000  
-USER=US09729264 @CGN.1.1.99 @runat.17092004.155107.2371 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	10	2.6	390	2	T03241	G-box binding fact
C 2	10	2.6	526	2	A56573	nuclear pore compl
C 3	10	2.6	2176	2	T13806	toucan gene protei
C 4	9	2.4	126	2	T52525	tyrosinase co-fact
C 5	9	2.3	165	1	CUOH	plastocyanin precu
C 6	9	2.3	287	2	T49329	hypothetical prote
C 7	9	2.3	332	2	JC7310	protein-glutamine
C 8	9	2.3	406	2	JC2089	L-ascorbate peroxi
C 9	9	2.4	430	2	T12282	protein-glutamine
C 10	9	2.4	662	2	A51555	mucin FIM-C.1 - Af
C 11	9	2.3	719	2	S60771	developmental regu
C 12	8	2.1	144	2	F70895	probable PE protei
C 13	8	2.1	150	2	T12547	hypothetical prote
C 14	8	2.1	172	2	S35568	sex-determining pr

15	8	2.1	189	2	B33513	hypothetical prote
16	8	2.1	200	2	T30223	hypothetical prote
17	8	2.1	215	2	C72722	probable agmatinas
18	8	2.1	233	2	S11292	SA85-1.1 protein -
C 19	8	2.1	284	2	G75447	hypothetical prote
C 20	8	2.1	296	2	H87324	nitrate ABC transp
C 21	8	2.1	303	2	G70784	probable mmpS3 pro
C 22	8	2.1	309	2	S03172	rRNA N-glycosidase
C 23	8	2.1	304	2	S11859	rRNA N-glycosidase
C 24	8	2.1	320	2	T21452	hypothetical prote
25	8	2.1	331	2	B47236	zinc-finger protei
26	8	2.1	340	2	I51547	probable RNA-bindi
27	8	2.1	347	2	I51546	probable RNA-bindi
C 28	8	2.1	366	2	C96583	hypothetical prote
C 29	8	2.1	417	2	T36464	probable oxidoredu
C 30	8	2.1	421	2	S71331	L-ascorbate peroxi
C 31	8	2.1	438	2	C75587	probable chalcone
C 32	8	2.1	443	1	I38239	transcription fact
C 33	8	2.1	476	2	C54415	transcription fact
C 34	8	2.1	477	2	A47236	zinc-finger protei
C 35	8	2.1	483	2	S44550	hypothetical prote
C 36	8	2.1	489	2	T26069	hypothetical prote
37	8	2.1	548	2	JC4917	signal transducing
C 38	8	2.1	623	1	A48123	cell cycle regulat
C 39	8	2.1	633	1	DNXLPA	polyadenylate-bind
C 40	8	2.1	637	2	A72532	probable DNA-direc
C 41	8	2.1	653	2	D95848	probable elongatio
C 42	8	2.1	671	2	A35912	homoeotic protein o
43	8	2.1	745	2	T44549	hypothetical prote
44	8	2.1	840	2	F84213	hypothetical prote
C 45	8	2.1	896	2	S36326	clathrin assembly
C 46	8	2.1	901	2	A44825	phosphoprotein, sy
C 47	8	2.1	915	2	S36327	clathrin assembly
C 48	8	2.1	982	2	A53253	microtubule-associ
49	8	2.1	990	2	T03784	probable receptor
C 50	8	2.1	995	2	T32466	hypothetical prote
C 51	8	2.1	1028	2	A56038	DNA-binding protei
C 52	8	2.1	1069	2	T00043	BH-protocadherin-a
C 53	8	2.1	1072	2	T00041	BH-protocadherin P
54	8	2.1	1087	2	T30844	serine-repeat anti
C 55	8	2.1	1102	2	T39943	hypothetical prote
C 56	8	2.1	1152	2	A33183	microtubule-associ
57	8	2.1	1157	2	F97255	fusion of alpha-gl
58	8	2.1	1197	2	D86317	protein F15H18.21
C 59	8	2.1	1200	2	T00042	BH-protocadherin P
C 60	8	2.1	1239	2	G02750	DNA-directed DNA p
C 61	8	2.1	1246	2	T00826	hypothetical prote
C 62	8	2.1	1816	2	A84845	probable ABC trans
C 63	8	2.1	1891	2	T13594	hypothetical prote
64	8	2.1	1920	2	T13893	gene hindsight pro
65	8	2.1	2223	2	A47447	calcium channel pr
C 66	8	2.1	2241	2	S09811	hypothetical prote
C 67	8	2.1	4845	2	T31067	BIR repeat contain
C 68	7	1.8	40	2	H81592	tyrosine 3-monooxy
69	7	1.8	45	2	PN0589	tyrosine 3-monooxy
70	7	1.8	45	2	PN0590	tyrosine 3-monooxy
71	7	1.8	54	2	D31443	ovomucoid, third d
72	7	1.8	54	2	A61588	ovomucoid (PSTI-ty
73	7	1.8	58	2	S39537	inter-alpha-trypsi
C 74	7	1.8	66	2	H72738	hypothetical prote
75	7	1.8	68	2	A52226	cyl-L protein - E
76	7	1.8	71	2	S45608	light-harvesting p
77	7	1.8	71	2	T49130	valine-rich protei
78	7	1.8	81	2	B70966	hypothetical prote
C 79	7	1.8	81	2	B70966	hypothetical prote
C 80	7	1.8	81	2	AC3489	antifreeze protein
C 81	7	1.8	85	1	FDPL4W	antifreeze protein
C 82	7	1.8	85	2	S25837	hypothetical prote
83	7	1.8	85	2	S72724	adenine phosphorib
C 84	7	1.8	91	2	A22592	antifreeze protein
C 85	7	1.8	92	2	A69851	hypothetical prote
C 86	7	1.8	94	2	UN0660	heat shock protein
C 87	7	1.8	94	2	S32105	chaperonin groES -

88 7 1.8 96 2 T30722 hypothetical prote  
89 7 1.8 97 2 S02376 antifreeze protein  
90 7 1.8 98 2 A64638 hypothetical prote  
91 7 1.8 99 1 R6MXL2 ribosomal protein  
92 7 1.8 104 2 B69794 hypothetical prote  
93 7 1.8 105 2 F84830 hypothetical prote  
94 7 1.8 109 2 T02762 another specific pr  
95 7 1.8 110 2 T49809 hypothetical prote  
96 7 1.8 111 2 E95195 hypothetical prote  
97 7 1.8 116 2 E48338 hypothetical prote  
98 7 1.8 117 1 WMBP5B gene s protein - p  
99 7 1.8 120 2 S31112 Ig heavy chain - h  
100 7 1.8 120 2 JN0432 ribosomal protein  
101 7 1.8 121 2 A69044 hypothetical prote  
102 7 1.8 122 2 S33461 lipid transfer pro  
103 7 1.8 125 2 A53692 synapsin I - mouse  
104 7 1.8 126 1 RPXFJY repetitive protein  
105 7 1.8 127 2 G95269 hypothetical prote  
106 7 1.8 129 2 T36732 hypothetical prote  
107 7 1.8 130 2 S41571 blastocidin-S deam  
108 7 1.8 133 1 SXADP1 hexon-associated p  
109 7 1.8 134 2 E81133 hypothetical prote  
110 7 1.8 137 2 H87619 hypothetical prote  
111 7 1.8 137 2 H88064 protein tial.4 [i  
112 7 1.8 139 2 H81886 hypothetical prote  
113 7 1.8 139 2 H81812 duplicated hypothe  
114 7 1.8 139 2 AC3405 hypothetical prote  
115 7 1.8 140 2 JC4607 hydrophobin 1 prec  
116 7 1.8 141 1 JCL287 ribonuclease Sa (E  
117 7 1.8 141 2 A90130 hypothetical prote  
118 7 1.8 141 2 A10216 probable exported  
119 7 1.8 142 2 F86443 unknown protein [i  
120 7 1.8 144 1 SXAD12 hexon-associated p  
121 7 1.8 145 2 S20631 Ig kappa chain - h  
122 7 1.8 148 2 T20053 hypothetical prote  
123 7 1.8 149 2 F72677 hypothetical prote  
124 7 1.8 150 2 T12547 hypothetical prote  
125 7 1.8 150 2 F86299 hypothetical prote  
126 7 1.8 156 1 YNEC cyanate hydratase  
127 7 1.8 156 2 A99678 cyanate aminohydro  
128 7 1.8 156 2 E85528 cyanate aminohydro  
129 7 1.8 158 2 T00933 RNA-binding protei  
130 7 1.8 159 2 D72738 hypothetical prote  
131 7 1.8 161 2 AF0709 probable exported  
132 7 1.8 162 1 YAZQ51 antigen 5.1 precur  
133 7 1.8 162 2 A26769 antigen 5.1 precur  
134 7 1.8 162 2 C72684 hypothetical prote  
135 7 1.8 163 2 T36518 hypothetical prote

## ALIGNMENTS

RESULT 1  
T03241  
G-box binding factor 1A - rice  
C;Species: Oryza sativa (rice)  
C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000  
C;Accession: T03241  
R;Nantel, A.; Quatrano, R.S.  
J. Biol. Chem. 271, 31296-31305, 1996  
A;Title: Characterization of three rice basic/leucine zipper factors, including two inh  
A;Reference number: Z14856; MUID:97094901; PMID:8940135  
A;Accession: T03241  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-390 <NAN>  
A;Cross-references: EMBL:U04295; NID:G435941; PIDN:AAC49556.1; PID:G435942  
A;Experimental source: cultivar Nipponbare  
C;Genetics:  
A;Note: osZip-1a  
C;Superfamily: fava bean G-box-binding protein; fos/jun DNA-binding domain homology  
C;Keywords: DNA binding; leucine zipper; transcription regulation

Alignment Scores:  
Pred. No.: 0.35 Length: 390  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.63% Indels: 0  
DB: 2 Gaps: 0  
US-09-729-264-1 (1-1175) x T03241 (1-390)  
Qy 836 ACAACACGACGGCGGCGAGCAGCAGCG 807  
Db 121 ThrThrThrAlaAlaAlaAlaAlaAla 130  
RESULT 2  
A56573  
nuclear pore complex glycoprotein p62 - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 21-Jul-1995 #sequence\_revision 28-Jul-1995 #text\_change 05-Nov-1999  
C;Accession: A56573  
R;Cordes, V.; Waizenegger, I.; Krohne, G.  
Eur. J. Cell Biol. 55, 31-47, 1991  
A;Title: Nuclear pore complex glycoprotein p62 of Xenopus laevis and mouse: cDNA cloning  
A;Reference number: A56573; MUID:92007945; PMID:1915419  
A;Accession: A56573  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-526 <COR>  
A;Cross-references: GB:S59342; NID:G236260; PIDN:AAB19953.1; PID:G236261  
A;Note: sequence extracted from NCBI backbone (NCBIN:59342, NCBI:P:59343)  
C;Comment: The amino end of this protein contains O-linked N-acetylglucosamine addition  
C;Keywords: Glycoprotein; nuclear membrane; peripheral membrane protein

Alignment Scores:  
Pred. No.: 0.337 Length: 526  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.63% Indels: 0  
DB: 2 Gaps: 0  
US-09-729-264-1 (1-1175) x A56573 (1-526)

Qy 836 ACAACACGACGGCGGCGAGCAGCAGCG 807  
Db 281 ThrThrThrAlaAlaAlaAlaAlaAla 290

RESULT 3  
T13806  
toucan gene protein - fruit fly (Drosophila melanogaster)  
C;Species: Drosophila melanogaster  
C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C;Accession: T13806  
R;Grammont, M.; Dastugue, B.; Couderc, J.L.  
Development 124, 4911-4926, 1997  
A;Title: The Drosophila toucan (toc) gene is required in germline cells for somatic cell  
A;Reference number: Z17769; MUID:98090047; PMID:9362455  
A;Accession: T13806  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-2176 <GRA>  
A;Cross-references: EMBL:Y14157; NID:G2760521; PIDN:CAA74574.1; PID:G2760522  
C;Genetics:  
A;Gene: toc  
A;Cross-references: FlyBase:FBgn0015600

Alignment Scores:  
Pred. No.: 0.282 Length: 2176  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.63% Indels: 0  
DB: 2 Gaps: 0



US-09-729-264-1 (1-1175) x T13806 (1-2176)

QY 836 ACAACACGACGGCGCAGCAGCAGCG 807  
Db 967 ThrThrThrAlaAlaAlaAlaAla 976

## RESULT 4

T52525 tyrosinase co-factor MelC2 [imported] - Streptomyces galbus

C:Species: Streptomyces galbus  
C:Date: 24-Oct-2000 #sequence\_revision 24-Oct-2000 #text\_change 24-Oct-2000  
C:Accession: T52525

R:Wehmeier, U.F.; Brass, N.; Roesler, C.; Piepersberg, W.

A:Description: Cloning and characterization of the mel-operon from Streptomyces galbus

A:Reference number: Z26098

A:Accession: T52525

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-126 <WEH>

A:Cross-references: EMBL:X95705; PIDN:CRA65004.1

A:Experimental source: DSM40480

C:Genetics:

A:Gene: melC1

Alignment Scores:  
Pred. No.: 4.2 Length: 126  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 2.37%  
Indels: 0  
Gaps: 0  
DB: 0

US-09-729-264-1 (1-1175) x T52525 (1-126)

QY 860 ACAACGACGACGAGTTCAGCCACA 834

Db 11 ThrThrAlaAlaAlaValAlaAlaThr 19

## RESULT 5

CUQH plastocyanin precursor - white campion

C:Species: Silene pratensis, lychnis alba (white campion, evening lychnis)  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 20-Feb-1998  
C:Accession: A24404

R:Smeekens, S.; de Groot, M.; van Binsbergen, J.; Weisbeek, P.

Nature 317, 456-458, 1985

A:Title: Sequence of the precursor of the chloroplast thylakoid lumen protein plastocyanin

A:Reference number: A24404

A:Accession: A24404

A:Molecule type: DNA

A:Residues: 1-165 <SME>

A:Cross-references: GB:X02965

C:Superfamily: plastocyanin

C:Keywords: chloroplast; copper; electron transfer; metalloprotein

F:1-66/Domain: transit peptide (chloroplast) #status predicted <TNP>

F:67-165/Product: plastocyanin #status predicted <MAT>

F:103,150,153,158/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted

Alignment Scores:  
Pred. No.: 4.06 Length: 165  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 2.31%  
Indels: 0  
Gaps: 0  
DB: 0

US-09-729-264-1 (1-1175) x CUQH (1-165)

QY 829 GTTGTGTGGTCGACCTGCTGCC 855

Db 49 ValValAlaAlaAlaThrAlaAlaAla 57

## RESULT 6

T49329

hypothetical protein B13N20.140 [imported] - Neurospora crassa

C:Species: Neurospora crassa

C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 18-Aug-2000

C:Accession: T49329

R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022

A:Accession: T49329

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-287 <SCH>

A:Cross-references: EMBL:AL355925; GSPDB:GN00116; NCSP:B13N20.140

A:Experimental source: BAC clone B13N20; strain OR74A

C:Genetics:

A:Gene: NCSP:B13N20.140

A:Map position: 6

A:Introns: 38/3

C:Superfamily: Neurospora crassa hypothetical protein B13N20.140

Alignment Scores:

Pred. No.: 3.79 Length: 287  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 2.31%  
Indels: 0  
Gaps: 0  
DB: 0

US-09-729-264-1 (1-1175) x T49329 (1-287)

QY 808 GCTGTGCTGCTGCCGCGCTGTT 834

Db 92 AlaAlaAlaAlaAlaValValVal 100

## RESULT 7

JC7310

protein-glutamine gamma-glutamyltransferase (EC 2.3.2.13) - Escherichia coli

N:Alternate names: microbial transglutaminase

C:Species: Escherichia coli

C:Date: 08-Sep-2000 #sequence\_revision 08-Sep-2000 #text\_change 17-Nov-2000

C:Accession: JC7310

R:Yokoyama, K.; Nakamura, N.; Seguro, K.; Kubota, K.

Biosci. Biotechnol. Biochem. 64, 1263-1270, 2000

A:Title: Overproduction of microbial transglutaminase in Escherichia coli, in vitro refo

A:Reference number: JC7310

A:Accession: JC7310

A:Molecule type: DNA

A:Residues: 1-332 <YOK>

A:Experimental source: strain JMI09

A:Note: Cross-reference

C:Comment: This enzyme belongs to a family of enzymes that catalyzes acyl transfer betwe

in the formation of epsilon-(gamma-glutamyl) lysine cross-linkages. This enzyme is invo

C:Genetics:

A:Gene: mtg

C:Superfamily: protein-glutamine gamma-glutamyltransferase

C:Keywords: aminoacyltransferase

Alignment Scores:  
Pred. No.: 3.72 Length: 332  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 2.31%  
Indels: 0  
Gaps: 0  
DB: 0

US-09-729-264-1 (1-1175) x JC7310 (1-332)

QY 356 CCTTACCGTCCAGTTATGGGAGCCT 382

Db 20 ProTyrArgProSerTyrGlyArgAla 28

```

A;Reference number: Z17482
A;Accession: TL2286
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-379, 'D' <MIC2>
A;Cross-references: EMBL:AF069316; NID:g3202025; PID:g3202026
C;Superfamily: cytochrome-c peroxidase
C;Keywords: heme; iron; metalloprotein; oxidoreductase
F;1-379, 'D' /Product: L-ascorbate peroxidase, stromal splice form #status predicted <SF>
F;119/Active site: His (distal axial ligand) #status predicted
F;248/Binding site: heme iron (His) (proximal axial ligand) #status predicted
F;280.309/Active site: Trp, Asp #status predicted

Alignment Scores:
Pred. No.: 3, 6 Length: 430
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.37% Indels: 0
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x TL2282 (1-430)

QY 833 ACAACGACGCGCGCAGCAGCAGCG 807
Db 11 ThrThrAlaAlaAlaAlaAla 19

RESULT 10
A45155
mucin FIM-C.1 - African clawed frog (fragment)
C;Species: Xenopus laevis (African clawed frog)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 21-Jul-2000
C;Accession: A45155
R;Hauser, F.; Hoffmann, W.
J. Biol. Chem. 267, 24620-24624, 1992
A;Title: P-domains as shuffled cysteine-rich modules in integratory muscin C.1 (FIM-C.1)
A;Reference number: A45155; MUID:93077556; PMID:1447205
A;Accession: A45155
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-662 <HAU>
A;Cross-references: GB:L02115; NID:g214147; PIDN:AAA74725.1; PID:g951460
C;Superfamily: trefoil homology
F;162-202/Domain: trefoil homology <TRF1>
F;307-347/Domain: trefoil homology <TRF2>
F;354-394/Domain: trefoil homology <TRF3>
F;526-566/Domain: trefoil homology <TRF4>
F;573-613/Domain: trefoil homology <TRF5>
F;621-661/Domain: trefoil homology <TRF6>

Alignment Scores:
Pred. No.: 3, 41 Length: 662
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.37% Indels: 0
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x A45155 (1-662)

QY 860 ACAACGACGACGAGTTGCAGCCACA 834
Db 3 ThrThrAlaAlaAlaAlaAlaThr 11

RESULT 11
S60771
developmental regulator flba - Emeritella nidulans
C;Species: Emeritella nidulans, Aspergillus nidulans
C;Date: 27-Apr-1996 #sequence_revision 06-Sep-1996 #text_change 06-Sep-1996
C;Accession: S60771; S60772
R;Lee, B.N.; Adams, T.H.
Mol. Microbiol. 14, 323-334, 1994
A;Title: Overexpression of flba, an early regulator of Aspergillus asexual sporulation,

```

A:Reference number: S60771; MUID:95131754; PMID:7830576

A:Accession: S60771

A:Molecule type: DNA

A:Residues: 1-11 <LEB>

A:Cross-references: EMBL:L24395

A:Accession: S60772

A:Molecule type: mRNA

A:Residues: 12-719 <LEW>

A:Cross-references: EMBL:L24395

C:Genetics:

A:Gene: flbA

Alignment Scores:

Pred. No.:	3.37	Length:	719
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.31%	Indels:	0
DB:	2	Gaps:	0

US-09-729-264-1 (1-1175) x S60771 (1-719)

QY 808 GCTGCTGCTGCGCGCGTGGTGT 834

Db 39 AlaAlaAlaAlaAlaValVal 47

RESULT 12

F70895

probable PE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: F70895

R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Rajandream, M.A.; Rogers, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Nature 393, 537-544, 1998

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: F70895

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-144 <COL>

A:Cross-references: GB:AL021897; GB:AL123456; NID:g3256022; PIDN:CAA17204.1; PID:e125196

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PE

Alignment Scores:

Pred. No.:	43	Length:	144
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.11%	Indels:	0
DB:	2	Gaps:	0

US-09-729-264-1 (1-1175) x F70895 (1-144)

QY 854 GCAGCAGCGTGGCGGCACACACA 831

Db 30 AlaAlaAlaValAlaAlaThrThr 37

RESULT 13

T12547

hypothetical protein DKFP586E1621.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 23-Jul-1999

C:Accession: T12547

R:Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, June 1999

A:Reference number: Z17528

A:Accession: T12547

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-150 <OTT>

A:Cross-references: EMBL:AL080235

A:Experimental source: adult uterus; clone DKFP586E1621

C:Genetics:

A:Note: DKFP586E1621.1

Alignment Scores:

Pred. No.:	42.8	Length:	150
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.11%	Indels:	0
DB:	2	Gaps:	0

US-09-729-264-1 (1-1175) x T12547 (1-150)

QY 830 ACGACGGCGCAGCAGCAGCAGCG 807

Db 127 ThrThrAlaAlaAlaAlaAla 134

RESULT 14

S35568

sex-determining protein Sry - multimammate rat (Mastomys hildebrandtii)

C:Species: Mastomys hildebrandtii

C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 24-Sep-1999

C:Accession: S35568

R:Tucker, P.K.; Lundrigan, B.L.

Nature 364, 715-717, 1993

A>Title: Rapid evolution of the sex determining locus in Old World mice and rats.

A:Reference number: S35565; MUID:93361118; PMID:8355784

A:Accession: S35568

A:Molecule type: DNA

A:Residues: 1-172 <TUC>

A:Cross-references: GB:L29542; NID:g496161; PIDN:AAA40587.1; PID:g496162

C:Genetics:

A:Gene: Sry

A:Map position: Y

C:Superfamily: unassigned HMG box proteins; HMG box homology

C:Keywords: DNA binding

F12-77/Domain: HMG box homology <HMG1>

Alignment Scores:

Pred. No.:	42.1	Length:	172
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.11%	Indels:	0
DB:	2	Gaps:	0

US-09-729-264-1 (1-1175) x S35568 (1-172)

QY 830 ACGACGGCGCAGCAGCAGCAGCG 807

Db 144 ThrThrAlaAlaAlaAlaAla 151

RESULT 15

B33513

hypothetical protein A - human immunodeficiency virus type 1

C:Species: human immunodeficiency virus type 1, HIV-1

C>Date: 20-Dec-1989 #sequence\_revision 20-Dec-1989 #text\_change 18-Jun-1993

C:Accession: B33513

R:Larocca, D.; Chao, L.A.; Seto, M.H.; Bruck, T.K.

Biochem. Biophys. Res. Commun. 163, 1006-1013, 1989

A>Title: Human T-cell leukemia virus minus strand transcription in infected T-cells.

A:Reference number: A33513; MUID:89391952; PMID:2476979

A:Accession: B33513

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: genomic RNA

A:Residues: 1-189 <LAR>

Alignment Scores:

Pred. No.:	41.6	Length:	189
------------	------	---------	-----

Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: Gaps: 0

US-09-729-264-1 (1-1175) x B33513 (1-189)

QY 809 CTGCTGCTGCGCGCGTGTGTTG 832

Db 111 LeuLeuLeuLeuProProSerLeu 118

RESULT 16

T30223

hypothetical protein U - Streptomyces hygroscopicus

C:Species: Streptomyces hygroscopicus

C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000

C:Accession: T30223

R:Apuricio, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun

Gene 169, 9-16, 1996

A>Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg

A:Reference number: Z20782; MUID:96186896; PMID:8635756

A:Accession: T30223

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-200 <APA>

A:Cross-references: EMBL:X86780; NID:g987088; PID:g987097; PIDN:CAA60457.1

Alignment Scores:

Pred. No.: 41.3 Length: 200  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: Gaps: 0

US-09-729-264-1 (1-1175) x T30223 (1-200)

QY 814 GCTGCTGCCGCGCGTGTGTTG 837

Db 65 AlaAlaAlaValValValVal 72

RESULT 17

C72722

probable agmatinase APE0316 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999

C:Accession: C72722

R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999

A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr

A:Reference number: A72450; MUID:99310339; PMID:10382966

A:Accession: C72722

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-215 <RAW>

A:Cross-references: DDBJ:AP000059; NID:g5103911; PIDN:BAA79271.1; PID:d1043057; PID:g510

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE0316

Alignment Scores:

Pred. No.: 40.9 Length: 215  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: Gaps: 0

US-09-729-264-1 (1-1175) x C72722 (1-215)

QY 859 GTTGTTTCTGCTGAGAGAAAAA 882

Db 9 ValValSerAlaValGluGluLys 16

RESULT 18

S11292

SA85-1.1 protein - Trypanosoma cruzi (fragment)

C:Species: Trypanosoma cruzi

C>Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 11-Jan-2000

C:Accession: S11292

R:Kahn, S.; van Voorhis, W.C.; Eisen, H.

J. Exp. Med. 172, 589-597, 1990

A>Title: The major 85-kD surface antigen of the mammalian form of Trypanosoma cruzi is

A:Reference number: S11292; MUID:90324879; PMID:1695668

A:Accession: S11292

A>Status: preliminary; translation not shown

A:Molecule type: mRNA

A:Residues: 1-233 <KAH>

A:Cross-references: EMBL:X53545; NID:gl0652; PID:g829230

C:Superfamily: trypanastigote-specific surface antigen

Alignment Scores:

Pred. No.: 40.5 Length: 233  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: Gaps: 0

US-09-729-264-1 (1-1175) x S11292 (1-233)

QY 848 CTGCTGCCGTTGTGTTCTGCTG 871

Db 218 LeuLeuProLeuLeuPheLeuLeu 225

RESULT 19

G75447

hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000

C:Accession: G75447

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: G75447

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-284 <WHI>

A:Cross-references: GB:AE001953; GB:AE000513; NID:g6458740; PIDN:AAF10595.1; PID:g64587

A:Experimental source: strain R1

C:Genetics:

A:Gene: DRI015

A:Map position: 1

Alignment Scores:

Pred. No.: 39.5 Length: 284  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: Gaps: 0

US-09-729-264-1 (1-1175) x G75447 (1-284)

QY 857 ACGGACGACGACGTTGACGCCACA 834

Db 154 ThrAlaAlaValAlaAlaThr 161

RESULT 20

H87324

nitrate ABC transporter, permease protein CC0610 [imported] - Caulobacter crescentus



F:26-284/Domain: rRNA N-glycosidase homology <RNG>  
F:93-98/Region: RNA binding  
F:137-143/Region: RNA binding  
F:162-186/Domain: activation peptide #status predicted <ACT>  
F:187-274/Domain: rRNA N-glycosidase beta chain #status predicted <BCH>

## Alignment Scores:

Pred. No.:	39.2	Length:	304
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.11%	Indels:	0
DB:	2	Gaps:	0

US-09-729-264-1 (1-1175) x S11859 (1-304)

QY 830 ACAGCGGCGAGCAGCAGCAGCG 807

Db 285 ThrThrAlaAlaAlaAlaAla 292

## RESULT 24

T21452

hypothetical protein F27D4.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T21452

R:Wilkinson, J.  
submitted to the EMBL Data Library, September 1996

A:Reference number: Z19424

A:Accession: T21452

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-320 &lt;WII&gt;

A:Cross-references: EMBL:Z79695; PIDN:CAB01968.1; GSPDB:GNC00019; CESP:F27D4.2

A:Experimental source: clone F27D4

C:Genetics:

A:Gene: CESP:F27D4.2

A:Map position: 1

A:Introns: 209/3

## Alignment Scores:

Pred. No.:	38.9	Length:	320
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.11%	Indels:	0
DB:	2	Gaps:	0

US-09-729-264-1 (1-1175) x T21452 (1-320)

QY 854 GCAGCAGCGTGCAGCAGCAGCAACA 831

Db 297 AlaAlaAlaAlaAlaAlaThr 304

## RESULT 25

B47236

zinc-finger protein Fur-1 - golden hamster

C:Species: Mesocricetus auratus (golden hamster)

C:Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 05-Apr-1995  
C:Accession: B47236

R:Kennedy, G.C.; Rutter, W.J.

Proc. Natl. Acad. Sci. U.S.A. 89, 11498-11502, 1992

A:Title: Fur-1, a zinc-finger protein that binds to purine-rich sequences, transactivates  
A:Reference number: B47236; MUID:93087555; PMID:1454839

A:Accession: B47236

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-331 &lt;KEN&gt;

A:Experimental source: insulinoma cell line T

A:Note: sequence extracted from NCBI backbone (NCBIP:119832)

## Alignment Scores:

Pred. No.:	38.8	Length:	331
------------	------	---------	-----

Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.06%	Indels:	0
DB:	2	Gaps:	0

US-09-729-264-1 (1-1175) x B47236 (1-331)

QY 808 GCTGCTGCTGCTGCGCGTGGTT 831

Db 90 AlaAlaAlaAlaAlaAlaVal 97

## RESULT 26

I51547

probable RNA-binding protein nrip-1B - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 24-Sep-1999  
C:Accession: I51547; S27944

R:Richter, K.; Good, P.J.; Dawid, I.B.  
New Biol. 2, 556-565, 1990

A:Title: A developmentally regulated, nervous system-specific gene in Xenopus encodes a

A:Reference number: I51546; MUID:91208109; PMID:1708282

A:Accession: I51547

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-340 &lt;RIC&gt;

A:Cross-references: GB:M34895; NID:g214631; PIDN:AAA49920.1; PID:g214632

C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprote

F:21-87/Domain: ribonucleoprotein repeat homology &lt;RRM1&gt;

F:110-176/Domain: ribonucleoprotein repeat homology &lt;RRM2&gt;

## Alignment Scores:

Pred. No.:	38.6	Length:	340
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.06%	Indels:	0
DB:	2	Gaps:	0

US-09-729-264-1 (1-1175) x I51547 (1-340)

QY 808 GCTGCTGCTGCTGCGCGTGGTT 831

Db 274 AlaAlaAlaAlaAlaAlaVal 281

## RESULT 27

I51546

Probable RNA-binding protein nrip-1A - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 24-Sep-1999  
C:Accession: I51546; S27943

R:Richter, K.; Good, P.J.; Dawid, I.B.  
New Biol. 2, 556-565, 1990

A:Title: A developmentally regulated, nervous system-specific gene in Xenopus encodes a

A:Reference number: I51546; MUID:91208109; PMID:1708282

A:Accession: I51546

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-347 &lt;RIC&gt;

A:Cross-references: GB:M34894; NID:g214629; PIDN:AAA49919.1; PID:g214630

C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprote

F:21-87/Domain: ribonucleoprotein repeat homology &lt;RRM1&gt;

F:110-176/Domain: ribonucleoprotein repeat homology &lt;RRM2&gt;

## Alignment Scores:

Pred. No.:	38.5	Length:	347
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.06%	Indels:	0
DB:	2	Gaps:	0

US-09-729-264-1 (1-1175) x I51546 (1-347)

QY 808 GCTGCTGCTGCTGCGCGCTGCT 831  
|||||  
Db 274 AlAAlAAlAAlAAlAAlVal 281

## RESULT 28

C96583  
hypothetical protein F20D21.2 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: C96583  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: C96583  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-366 <STO>  
A:Cross-references: GB:AE005173; NID:94585964; PIDN:AAD25600.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F20D21.2  
A:Map position: 1

Alignment Scores:  
Pred. No.: 38.3 Length: 366  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x C96583 (1-366)

QY 830 ACAGCGGCGAGCGAGCAGCG 807  
|||||  
Db 203 ThrThrAlaAlaAlaAlaAla 210

## RESULT 29

T36464  
Probable oxidoreductase - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T36464  
R:Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, September 1999  
A:Reference number: Z21607  
A:Accession: T36464  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-417 <SEE>  
A:Cross-references: EMBL:AL110470; PIDN:CAB54161.1; GSPDB:GN00070; SCOEDB:SCF85.04  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SCF85.04

Alignment Scores:  
Pred. No.: 37.7 Length: 417  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x T36464 (1-417)

QY 854 GCAGCAGAGTTGCAGCCACAACA 831  
|||||  
Db 371 AlAAlAAlAAlAAlAAlThrThr 378

## RESULT 30

S71331  
L-ascorbate peroxidase (EC 1.11.1.11) precursor - spinach (fragment)  
C:Species: Spinacia oleracea (spinach)  
C>Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Jul-2000  
C:Accession: S71331; S71329; S71330  
R:Shigeoka, S.  
submitted to the EMBL Data Library, October 1995  
A:Reference number: S71331  
A:Accession: S71331  
A:Molecule type: mRNA  
A:Residues: 1-421 <SHI>  
A:Cross-references: EMBL:D77997  
R:Shikawa, T.; Sakai, K.; Yoshimura, K.; Takeda, T.; Shigeoka, S.  
FEBS Lett. 384, 289-293, 1996  
A:Title: cDNAs encoding spinach stromal and thylakoid-bound ascorbate peroxidase, differ  
A:Reference number: S71329; MUID:96197808; PMID:8617374  
A:Accession: S71329

A:Molecule type: mRNA

A:Residues: 7-421 <ISH1>

A:Cross-references: EMBL:D77997; NID:91944506; PIDN:BAAL9611.1; PID:G1944507

A:Accession: S71330

A:Molecule type: mRNA

A:Residues: 7-370, 'D' <ISH2>

A:Cross-references: EMBL:D83669; NID:91944508; PIDN:BAAL2039.1; PID:G1369920

C:Genetics:

A:Genome: nuclear

C:Superfamily: cytochrome-c peroxidase

C:Keywords: chloroplast; heme; iron; metalloprotein; oxidoreductase

F:1-76/Domain: transit peptide (chloroplast) #status predicted <TNP>

F:77-421/Product: ascorbate peroxidase #status predicted <WAT>

F:110/Active site: His (distal axial ligand) #status predicted

F:239/Binding site: heme iron (His) (proximal axial ligand) #status predicted

F:271,300/Active site: Trp, Asp #status predicted

Alignment Scores:

Pred. No.: 37.6 Length: 421  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x S71331 (1-421)

QY 836 ACAACAAGCAGCGCGCAGCA 813  
|||||  
Db 11 ThrThrThrAlaAlaAlaAla 18

## RESULT 31

C75587  
probable chalcone synthase - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 12-Jun-2003  
C:Accession: C75587  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: AY5250; MUID:20036896; PMID:10567266  
A:Accession: C75587  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-438 <WHI>  
A:Cross-references: GB:AE001863; GB:AE001825; NID:96460670; PIDN:AAF12453.1; PID:G646074  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DRA0326

Alignment Scores:

Pred. No.: 37.7 Length: 417  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 2 Gaps: 0

A;Map position: 2

C;Superfamily: chalcone/stilbene synthase

Alignment Scores:  
Pred. No.: 37.4 Length: 438  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x C75587 (1-438)

Qy 847 GCTGCTGCGTGTGTGTTCTGCT 870

Db 270 AlaAlaAlaValValSerAla 277

#### RESULT 32

I38239  
transcription factor SOX3 - human  
N;Alternate names: SRY (sex determining region Y)-box 3  
C;Species: Homo sapiens (man)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 28-Jan-2000  
C;Accession: I38239; I38242; S67816  
R;Stevanovic, M.; Lovell-Badge, R.; Collignon, J.; Goodfellow, P.N.  
Hum. Mol. Genet. 2, 2013-2018, 1993  
A;Title: SOX3 is an X-linked gene related to SRY.  
A;Reference number: I38239; MUID:94154672; PMID:8111369  
A;Accession: I38239  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-443 <STE1>  
A;Cross-references: EMBL:X71135; NID:9468790; PIDN:CAA50465.1; PID:9530020  
A;Accession: I38242  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 143-158, 'P', 160-218 <STE2>  
A;Cross-references: EMBL:X71137; NID:9468793; PIDN:CAA50467.1; PID:9468794  
C;Genetics:  
A;Gene: GDB:SOX3; SOX-3; SOXB  
A;Cross-references: GDB:250376; OMIM:313430  
A;Map position: Xq26-Xq27  
A;Superfamily: human SOX3 protein; HMG box homology  
F;136-211/Domain: HMG box homology <HMG>

Alignment Scores:  
Pred. No.: 37.4 Length: 443  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x I38239 (1-443)

Qy 827 ACGCGGCAGCAGCAGCGCTAT 804

Db 338 ThrAlaAlaAlaAlaAlaTyr 345

#### RESULT 33

C54415  
transcription factor HBP-1b(c1) - wheat (fragment)  
C;Species: Triticum aestivum (common wheat)  
C;Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 21-Jul-2000  
C;Accession: C54415  
R;Mikami, K.; Sakamoto, A.; Iwabuchi, M.  
J. Biol. Chem. 269, 9974-9985, 1994  
A;Title: The HBP-1 family of wheat basic/leucine zipper proteins interacts with overlapping  
A;Reference number: A54415; MUID:94193693; PMID:8144592  
A;Accession: C54415  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-476 <MIK>

A;Cross-references: GB:D12921; NID:95926681; PIDN:BAAO2305.2; PID:95926682  
C;Superfamily: fos/jun DNA-binding domain homology  
F;184-227/Domain: fos/jun DNA-binding domain homology <FJD>

Alignment Scores:  
Pred. No.: 37 Length: 476  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x C54415 (1-476)

Qy 830 ACGACGGCGCAGCAGCAGCG 807

Db 101 ThrThrAlaAlaAlaAlaAla 108

#### RESULT 34

A47236  
zinc-finger protein Pur-1 - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Apr-1995  
C;Accession: A47236  
R;Kennedy, G.C.; Rutter, W.J.  
Proc. Natl. Acad. Sci. U.S.A. 89, 11498-11502, 1992  
A;Title: Pur-1, a zinc-finger protein that binds to purine-rich sequences, transactivates  
A;Reference number: A47236; MUID:93087555; PMID:1454839  
A;Accession: A47236  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: nucleic acid  
A;Residues: 1-477 <KEN>  
A;Experimental source: islet cell line beta TC3  
A;Note: sequence extracted from NCBI backbone (NCBIP:119831)

Alignment Scores:  
Pred. No.: 37 Length: 477  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x A47236 (1-477)

Qy 808 GCTGCTGCTGCGCGTCTGTT 831

Db 103 AlaAlaAlaAlaAlaValVal 110

#### RESULT 35

S44550  
hypothetical protein YBR288c - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein YBR2035  
C;Species: Saccharomyces cerevisiae  
C;Date: 08-Jun-1994 #sequence\_revision 09-Sep-1994 #text\_change 29-Oct-1999  
C;Accession: S44550; S46170; S39144  
R;Holmstrom, K.; Brandt, T.; Kalliesoe, T.  
Yeast 10(Suppl.A), S47-S62, 1994  
A;Title: The sequence of a 32420 bp segment located on the right arm of chromosome II  
A;Reference number: S44537; MUID:94378722; PMID:8091861  
A;Accession: S44550  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 1-483 <HOL>  
A;Cross-references: EMBL:X76053; NID:9600025; PIDN:CAA53651.1; PID:9429133  
R;Brandt, T.; Christiansen, C.; Holmstrom, K.; Kalliesoe, T.  
submitted to the Protein Sequence Database, August 1994  
A;Reference number: S46157  
A;Accession: S46170  
A;Molecule type: DNA  
A;Residues: 1-483 <BRA>  
A;Cross-references: EMBL:Z36157; NID:9536739; PIDN:CAA85253.1; PID:9536740; MIPS:YBR288  
C;Genetics:



A:Gene: SGD:APM3  
A:Cross-references: SGD:S0000492; MIPS:YBR288C  
A:Map position: 2R

Alignment Scores:  
Pred. No.: 37 Length: 483  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x S44550 (1-483)

QY 862 CAACAACGCGACGACGCTGCGAG 839

Db 174 GInGInAArgGInGInGInleuGIn 181

RESULT 36

T26069

hypothetical protein W02A11.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Sep-2000

C:Accession: T26069

R:White, S.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z20147

A:Accession: T26069

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-489 <WIL>

A:Cross-references: EMBL:Z82062; PIDN:CAB04890.1; GSPDB:GN00019; CESP:W02A11.3

A:Experimental source: clone W02A11

C:Genetics:

A:Gene: CESP:W02A11.3

A:Map position: 1

A:Introns: 58/3; 129/3; 361/3; 444/3

C:Superfamily: RING finger homology

F:429-479/Domain: RING finger homology <RRN>

Alignment Scores:

Pred. No.: 36.9 Length: 489  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x T26069 (1-489)

QY 808 GCTGCTGCTGCGCGCGCTGCTT 831

Db 309 AlaAlaAlaAlaAlaAlaVal 316

RESULT 37

JC4917

signal transducing adaptor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 26-Sep-1996 #sequence\_revision 01-Nov-1996 #text\_change 21-Jul-2000

C:Accession: JC4917

R:Takehita, T.; Arita, T.; Asao, H.; Tanaka, N.; Higuchi, M.; Kuroda, H.; Kaneko, K.; M

Biochem. Biophys. Res. Commun. 225, 1035-1039, 1996

A:Title: Cloning of a novel signal-transducing adaptor molecule containing an SH3 domain

A:Reference number: JC4916; MUID:96374438; PMID:8780729

A:Accession: JC4917

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-548 <TKA>

A:Cross-references: GB:U43900; NID:g1556460; PIDN:AA052840.1; PID:g3645912

A:Experimental source: T cell

C:Comment: This protein contains a Src-homology 3 domain and the immunoreceptor tyrosine

ne receptors.

C:Genetics:

A:Map position: 2A2-B  
C:Superfamily: SH3 homology  
F:2-540/Product: signal transducing adaptor molecule #status predicted <MAT>  
F:217-264/Domain: SH3 homology <SH3>  
F:359-387/Region: immunoreceptor tyrosine-based activation motif

Alignment Scores:  
Pred. No.: 36.4 Length: 548  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x JC4917 (1-548)

QY 808 GCTGCTGCTGCTGCGCGCTGCTT 831

Db 488 AlaAlaAlaAlaAlaAlaVal 495

RESULT 38

A48123

cell cycle regulatory protein HPC2 - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YBR1503; protein YBR215W

C:Species: Saccharomyces cerevisiae

C:Date: 07-May-1993 #sequence\_revision 03-May-1994 #text\_change 29-Oct-1999

C:Accession: A48123; S46091; S46089; S27426

R:Xu, H.; Kim, U.J.; Schuster, T.; Grunstein, M.

Mol. Cell. Biol. 12, 5249-5259, 1992

A:Title: Identification of a new set of cell cycle-regulatory genes that regulate S-phases

A:Reference number: A48123; MUID:93024471; PMID:1405694

A:Accession: A48123

A:Molecule type: DNA

A:Residues: 1-623 <XUL>

A:Cross-references: EMBL:M94207; NID:gl71699; PIDN:AAA34684.1; PID:gl71700

A>Note: sequence extracted from NCBI backbone (NCBIN:115647, NCBI:115649)

R:Dubois, E.; El Bakoury, M.; Glansdorff, N.; Messenguy, F.; Pierard, A.; Scherens, B.;

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45782

A:Accession: S46091

A:Molecule type: DNA

A:Residues: 1-623 <DUB>

A:Cross-references: EMBL:Z36084; NID:g536601; PIDN:CAA85179.1; PID:g536602; MIPS:YBR215W

R:Rieger, M.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45734

A:Accession: S46089

A:Molecule type: DNA

A:Residues: 1-261 <RIE>

A:Cross-references: EMBL:Z36084; MIPS:YBR215W

A:Gene: SGD:HPC2

A:Cross-references: SGD:S0000419; MIPS:YBR215W

A:Map position: 2R

C:Keywords: transcription regulation

Alignment Scores:

Pred. No.: 35.8 Length: 623  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x A48123 (1-623)

QY 706 GATAATACACCTGGAATATA 683

Db 251 AspAsnThrProGlyIleIle 258

RESULT 39

DNXLPA

polyadenylate-binding protein - African clawed frog

N;Alternate names: poly(A)-binding protein  
C;Species: Xenopus laevis (African clawed frog)  
C;Date: 27-Feb-1990 #sequence\_revision 26-May-1994 #text\_change 22-Jun-1999  
C;Accession: A32323; S12000  
R;Zelus, B.D.; Giebelhaus, D.H.; Bib, D.W.; Kenner, K.A.; Moon, R.T.  
Mol. Cell. Biol. 9, 2756-2760, 1989  
A;Title: Expression of the poly(A)-binding protein during development of Xenopus laevis.  
A;Reference number: A32323; MUID:89343997; PMID:2761544  
A;Accession: A32323  
A;Molecule type: mRNA  
A;Residues: 1-633 <ZEL>  
A;Cross-references: GB:M27072; NID:G623597; PIDN:AAA60936.1; PID:G623598  
A;Note: the authors translated the codon AAT for residue 197 as Phe, TTT for residue for  
G for residue 563 as Pro  
R;Nietfeld, W.; Mentzel, H.; Pieler, T.  
EMBO J. 9, 3699-3705, 1990  
A;Title: The Xenopus laevis poly(A) binding protein is composed of multiple functionally  
A;Reference number: S12000; MUID:91006071; PMID:2209558  
A;Accession: S12000  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-251, 'N', 253-283, 'K', 285-429, 'S', 431-602, 'S', 604-633 <NIE>  
A;Cross-references: GB:X57483; NID:G64969; PIDN:CAA40721.1; PID:G64570  
C;Superfamily: polyadenylate-binding protein; ribonucleoprotein repeat homology  
C;Keywords: duplication; nucleus; RNA binding  
F;12-79/Domain: ribonucleoprotein repeat homology <RRM1>  
F;13-18/Region: RNA-binding RNP2 motif  
F;52-59/Region: RNA-binding RNP1 motif  
F;100-165/Domain: ribonucleoprotein repeat homology <RRM2>  
F;101-106/Region: RNA-binding RNP2 motif  
F;138-145/Region: RNA-binding RNP1 motif  
F;192-258/Domain: ribonucleoprotein repeat homology <RRM3>  
F;193-198/Region: RNA-binding RNP2 motif  
F;231-238/Region: RNA-binding RNP1 motif  
F;295-360/Domain: ribonucleoprotein repeat homology <RRM4>  
F;296-301/Region: RNA-binding RNP2 motif  
F;333-340/Region: RNA-binding RNP1 motif

Alignment Scores:  
Pred. No.: 35.7 Length: 633  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x DNXLPA (1-633)  
QY 830 ACGACGCGCAGCAGCAGCG 807  
Db 494 ThrThrAlaAlaAlaAlaAla 501

RESULT 40  
A72532  
Probable DNA-directed DNA polymerase APE2229 - Aeropyrum pernix (strain K1)  
C;Species: Aeropyrum pernix  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C;Accession: A72532  
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A;Reference number: A72450; MUID:99310339; PMID:10382966  
A;Accession: A72532  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-637 <KAW>  
A;Cross-references: DDBJ:AP000063; NID:G5105654; PIDN:BAAB1241.1; PID:dl045027; PID:G510  
A;Experimental source: strain K1  
C;Genetics:  
A;Gene: APE2229

Alignment Scores:  
Pred. No.: 35.7 Length: 633  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x D95848 (1-653)  
QY 122 GCGAGCCTGGAGCCCTTCAGGAC 99  
Db 304 AlaSerLeuGlyAlaLeuGlnAse 311

RESULT 42  
A35912  
homeotic protein orthodenticle - fruit fly (Drosophila melanogaster)  
N;Alternate names: homeotic protein otd  
C;Species: Drosophila melanogaster  
C;Date: 08-Mar-1991 #sequence\_revision 08-Mar-1991 #text\_change 24-Sep-1999  
C;Accession: A35912; S18200  
R;Finkelstein, R.; Smouse, D.; Capaci, T.M.; Spradling, A.C.; Perrimon, N.  
Genes Dev. 4, 1516-1527, 1990  
A;Title: The orthodenticle gene encodes a novel homeo domain protein involved in the d  
A;Reference number: A35912; MUID:91071580; PMID:1979296  
A;Accession: A35912  
A;Molecule type: mRNA  
A;Residues: 1-671 <FIN>  
A;Cross-references: GB:X58983; NID:G8311; PIDN:CAA41732.1; PID:G8312  
C;Genetics:  
A;Gene: orthodenticle

Pred. No.: 35.7 Length: 637  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x A72532 (1-637)  
QY 449 ACACCTGCACCGCTCCCGGTAT 472  
Db 335 ThrLeuAspProAlaProGlyTyr 342

RESULT 41  
D95848  
probable elongation factor G protein [imported] - Sinorhizobium meliloti (strain 1021)  
C;Species: Sinorhizobium meliloti  
C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C;Accession: D95848  
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Herna  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A;Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing end  
A;Reference number: A95842; MUID:21396508; PMID:11481431  
A;Accession: D95848  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-653 <KUR>  
A;Cross-references: GB:AL591985; PIDN:CAC48452.1; PID:gl5139924; GSPDB:GN00167  
A;Experimental source: strain 1021, megaplasmid pSymb  
R;Galibert, F.; Finan, T.M.; Long, S.K.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler  
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure  
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K  
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A;Reference number: A96039; MUID:21368234; PMID:11474104  
A;Contents: annotation  
C;Genetics:  
A;Gene: fuaA2; SMB20049  
A;Genome: plasmid

Alignment Scores:  
Pred. No.: 35.6 Length: 653  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x D95848 (1-653)  
QY 122 GCGAGCCTGGAGCCCTTCAGGAC 99  
Db 304 AlaSerLeuGlyAlaLeuGlnAse 311

RESULT 42  
A35912  
homeotic protein orthodenticle - fruit fly (Drosophila melanogaster)  
N;Alternate names: homeotic protein otd  
C;Species: Drosophila melanogaster  
C;Date: 08-Mar-1991 #sequence\_revision 08-Mar-1991 #text\_change 24-Sep-1999  
C;Accession: A35912; S18200  
R;Finkelstein, R.; Smouse, D.; Capaci, T.M.; Spradling, A.C.; Perrimon, N.  
Genes Dev. 4, 1516-1527, 1990  
A;Title: The orthodenticle gene encodes a novel homeo domain protein involved in the d  
A;Reference number: A35912; MUID:91071580; PMID:1979296  
A;Accession: A35912  
A;Molecule type: mRNA  
A;Residues: 1-671 <FIN>  
A;Cross-references: GB:X58983; NID:G8311; PIDN:CAA41732.1; PID:G8312  
C;Genetics:  
A;Gene: orthodenticle

A;Cross-references: FlyBase:FBgn0004102  
A;Map position: X 7F1-8A5  
C;Superfamily: unassigned homeobox proteins; homeobox homology  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:74-130/Domain: homeobox homology <HOX>

Alignment Scores:  
Pred. No.: 35,5 Length: 671  
Score: 8,00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x A35912 (1-671)

QY 851 GCAGCAGTTGCAGCCACACACG 828  
Db 574 AlaAlaValAlaAlaThrThrThr 581

RESULT 43  
T44549  
hypotheical protein PA0625 [imported] - Pseudomonas aeruginosa  
C;Species: Pseudomonas aeruginosa  
C;Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 31-Dec-2000  
C;Accession: T44549; G83567  
R;Nakayama, K.; Takashima, K.; Ishihara, H.; Shinomiya, T.; Kageyama, M.; Kanaya, S.; Oh  
submitted to the EMBL Data Library, August 1999  
A;Description: Genetic relationship between bacteriocins and bacteriophages.  
A;Reference number: Z27290  
A;Accession: T44549  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-745 <NA>  
A;Cross-references: EMBL:AB030825; PIDN:BA083164.1  
A;Experimental source: strain PA01  
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A;Reference number: A82950; MUID:20437337; PMID:10984043  
A;Accession: G83567  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-745 <STO>  
A;Cross-references: GB:AE004498; GB:AE004091; NID:g9946491; PIDN:AAG04014.1; GSPDB:GN001  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: PA0625

Alignment Scores:  
Pred. No.: 35 Length: 745  
Score: 8,00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x T44549 (1-745)

QY 708 AGTTTACCGAGTTAGGTTTTC 731  
Db 554 SerLeuProSerLeuGlyPheSer 561

RESULT 44  
F84213  
hypotheical protein Vng0555c [imported] - Halobacterium sp. NRC-1  
C;Species: Halobacterium sp. NRC-1  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: F84213  
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S  
; Leitthauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl

Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
A;Title: Genome sequence of Halobacterium species NRC-1.  
A;Reference number: A84160; MUID:20504483; PMID:11016950  
A;Accession: F84213  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-840 <STO>  
A;Cross-references: GB:AE004437; NID:gi0580152; PIDN:ARG19074.1; GSPDB:GN00138  
C;Genetics:  
A;Gene: VNG0555C

Alignment Scores:  
Pred. No.: 34,5 Length: 840  
Score: 8,00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x F84213 (1-840)

QY 811 GCTGCTGCTGCGCGCTGCTGTT 834  
Db 238 AlaAlaAlaAlaAlaValVal 245

RESULT 45  
S36326  
clathrin assembly protein AP180 short form - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 08-Oct-1999  
R;Morris, S.A.; Schroeder, S.; Plessmann, U.; Weber, K.; Ungewickell, E.  
EMBO J. 12, 667-675, 1993  
A;Title: Clathrin assembly protein AP180: primary structure, domain organization and ide  
A;Reference number: S36326; MUID:93178442; PMID:8440257  
A;Accession: S36326  
A;Molecule type: mRNA  
A;Residues: 1-896 <MOR>  
A;Cross-references: EMBL:X68877; NID:g55724; PIDN:CAA48748.1; PID:g55725  
C;Keywords: clathrin binding

Alignment Scores:  
Pred. No.: 34,2 Length: 896  
Score: 8,00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x S36326 (1-896)

QY 833 ACAACGACGGCGCAGCAGCA 810  
Db 532 ThrThrAlaAlaAlaAlaAla 539

Search completed: September 18, 2004, 23:03:08  
Job time : 38.1682 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 18, 2004, 22:38:31 ; Search time 12.1337 Seconds  
(without alignments)  
10084.722 Million cell updates/sec

Title: US-09-729-264-1  
Perfect score: 389  
Sequence: 1 ctgtctcccatctgaataa.....gtaatacaactgtagtatag 1175

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Word size: 0

Total number of hits satisfying chosen parameters: 280262

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Listing first 135 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2.1/USPTO\_spool\_p/US09729264/runat.17092004.155106.2344/app\_query.fasta.1.4117  
-DB=SwissProt 42 -QFMT=fastan -SUFFIX=oli.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=135 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=0 -ALIGN=45 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09729264@cgn.1.1.53 @runat.17092004.155106.2344 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	10	2.6	526	1 NU62 MOUSE	Q63850 mus musculus
C 2	10	2.6	1556	1 PROS_DROVI	Q9U6A1 drosophila
C 3	9	2.4	126	1 TYRT_STRAL	P55046 streptomyces
C 4	9	2.3	165	1 PLAS_SILPR	P07030 silene prat
C 5	9	2.3	407	1 TGAS_STRMB	P81453 streptomyces
C 6	9	2.4	662	1 MUC1_XENLA	Q05049 xenopus lae
C 7	9	2.3	719	1 FLBA_EMENI	P38093 emericella
C 8	8	2.1	68	1 BRH2_HUMAN	Q9NY43 homo sapien
C 9	8	2.1	105	1 RL22_LEIDO	O43940 leishmania
C 10	8	2.1	146	1 SSB_DROME	P54622 drosophila
C 11	8	2.1	158	1 HUNE_DROMM	O46248 drosophila
C 12	8	2.1	184	1 YC17_DROME	Q9VH95 drosophila
C 13	8	2.1	299	1 WMS3_MYCTU	Q10390 mycobacteri
C 14	8	2.1	303	1 ALB3_MAIZE	P10593 zea mays (m
C 15	8	2.1	304	1 RIP9_MAIZE	P25892 zea mays (m
C 16	8	2.1	309	1 SOX3_XENLA	P55863 xenopus lae
C 17	8	2.1	319	1 MGB2_HUMAN	O15479 homo sapien
C 18	8	2.1	331	1 MAZ_MESAU	P56670 mesocricetu

C 19	8	2.1	361	1 IE63_PRVKA	Q85232 pseudorabie
C 20	8	2.1	384	1 BRH2_RAT	Q88181 rattus norv
C 21	8	2.1	393	1 SETB_HUMAN	Q9NTR1 homo sapien
C 22	8	2.1	403	1 FXQ1_HUMAN	Q3C009 homo sapien
C 23	8	2.1	415	1 RL3_DROME	O16797 drosophila
C 24	8	2.1	443	1 SOX3_HUMAN	P41225 homo sapien
C 25	8	2.1	476	1 HBB2_WHEAT	Q41558 triticum ae
C 26	8	2.1	477	1 MAZ_MOUSE	P56671 mus musculu
C 27	8	2.1	483	1 APM3_YEAST	P38153 saccharomyc
C 28	8	2.1	485	1 MKR4_HUMAN	Q13434 homo sapien
C 29	8	2.1	606	1 ECR_BOMMO	P49881 bombyx mori
C 30	8	2.1	623	1 HPC2_YEAST	Q01448 saccharomyc
C 31	8	2.1	633	1 PAB1_XENLA	P20965 xenopus lae
C 32	8	2.1	671	1 HMOC_DROME	P22810 drosophila
C 33	8	2.1	752	1 R511_TRYCR	P18269 trypanosoma
C 34	8	2.1	770	1 TLE1_HUMAN	Q04724 homo sapien
C 35	8	2.1	770	1 TLE1_MOUSE	Q62440 mus musculu
C 36	8	2.1	881	1 PRP2_HUMAN	Q99595 homo sapien
C 37	8	2.1	901	1 A180_MOUSE	Q61548 mus musculu
C 38	8	2.1	915	1 A180_RAT	Q05140 rattus norv
C 39	8	2.1	1028	1 OVO_DROME	P51521 drosophila
C 40	8	2.1	1069	1 PCH7_HUMAN	O60245 homo sapien
C 41	8	2.1	1102	1 YG49_SCHPO	O60184 schizosacch
C 42	8	2.1	1152	1 MAP4_HUMAN	P27816 homo sapien
C 43	8	2.1	1239	1 DFG1_HUMAN	P54098 homo sapien
C 44	8	2.1	2223	1 CCAE_DISOM	P56699 discopyge o
C 45	8	2.1	2241	1 TEGU_HCMVA	P16785 human cytom
C 46	8	2.1	4829	1 BIR6_HUMAN	Q9NR09 homo sapien
C 47	7	1.8	54	1 IOVO_AEPAR	P52241 aepyrodus
C 48	7	1.8	54	1 IOVO_LEIOC	P05581 leiposa ocel
C 49	7	1.8	74	1 SR14_MACRA	O18881 macaca radi
C 50	7	1.8	85	1 ANP4_PSEAM	P02734 pseudopleur
C 51	7	1.8	89	1 RS18_THETN	Q816m3 thermoanaer
C 52	7	1.8	91	1 ANPX_PSEAM	P07835 pseudopleur
C 53	7	1.8	91	1 ANPY_PSEAM	P23699 pseudopleur
C 54	7	1.8	92	1 YU1A_BACSU	O34679 bacillus su
C 55	7	1.8	94	1 CH10_LACIA	P37283 lactococcus
C 56	7	1.8	94	1 ANP1_LACLC	Q9AEP8 lactococcus
C 57	7	1.8	97	1 ANP1_LIMFE	P09031 limanda fer
C 58	7	1.8	98	1 Y945_HELEP	P08015 araneus dia
C 59	7	1.8	99	1 RL12_METVA	P31855 pseudomonas
C 60	7	1.8	101	1 RL12_METTL	Q889X9 pseudomonas
C 61	7	1.8	105	1 RL22_LEIDO	Q43194 sorghum bic
C 62	7	1.8	109	1 ZM33_MAIZE	O52179 chromatium
C 63	7	1.8	117	1 VPS_BPRPD	P80517 araneus dia
C 64	7	1.8	120	1 CUL5_ARADI	Q9GXU7 mus musculu
C 65	7	1.8	120	1 RL7_PSEPK	P78986 aspergillus
C 66	7	1.8	121	1 RL7_PSESM	P48312 human adeno
C 67	7	1.8	122	1 NLT2_SORBI	P32539 human adeno
C 68	7	1.8	124	1 SGP2_CHRVI	P52748 pisolithus
C 69	7	1.8	126	1 CU24_ARADI	P30289 streptomyce
C 70	7	1.8	127	1 CU26_ARADI	P59775 chlamydomon
C 71	7	1.8	128	1 PRK2_MOUSE	P03284 human adeno
C 72	7	1.8	130	1 BSD_ASPTE	Q8VMUL mycoplasma
C 73	7	1.8	132	1 HEX9_ADE40	Q8VNU1 human adeno
C 74	7	1.8	133	1 HEX9_ADE41	P32539 human adeno
C 75	7	1.8	140	1 RNP1_PISTU	P52748 pisolithus
C 76	7	1.8	141	1 RNS3_STRAU	P30289 streptomyce
C 77	7	1.8	141	1 RRB2_CHLRE	P59775 chlamydomon
C 78	7	1.8	144	1 HEX9_ADE12	P03284 human adeno
C 79	7	1.8	155	1 RS7_MYCMY	Q8VMUL mycoplasma
C 80	7	1.8	156	1 CYNS_ECO57	Q8VNU1 human adeno
C 81	7	1.8	156	1 CYNS_ECOLI	P58704 escherichia
C 82	7	1.8	161	1 KNL2_BOMOR	P00816 escherichia
C 83	7	1.8	162	1 EXP1_PLAFA	P83059 bombina ori
C 84	7	1.8	164	1 GLB1_CHLEU	P04926 plasmodium
C 85	7	1.8	164	1 TIG3_HUMAN	Q08753 chlamydomon
C 86	7	1.8	168	1 SULA_SERMA	Q9UL19 homo sapien
C 87	7	1.8	170	1 LSPA_PSEFL	P08845 serratia ma
C 88	7	1.8	171	1 GLB2_CHLEU	P17942 pseudomonas
C 89	7	1.8	171	1 LSPA_PSEPK	P52334 chlamydomon
C 90	7	1.8	173	1 LSPA_PSESM	Q88991 pseudomonas
C 91	7	1.8	175	1 AXIS_ARATH	Q88993 pseudomonas
C 92	7	1.8	175	1 AXIS_ARATH	Q9XFM0 arabidopsis

C 92	7	1.8	180	1	RL10_TREPA	083267 treponema p
C 93	7	1.8	181	1	RRP3_HORVU	048609 hordeum vul
C 94	7	1.8	181	1	RRP3_HORVU	048609 hordeum vul
C 95	7	1.8	182	1	K2C3_BOVIN	P04261 bos taurus
C 96	7	1.8	185	1	LEP1_BACAM	P41026 bacillus am
C 97	7	1.8	186	1	VALD_TRYEB	P17962 trypanosoma
C 98	7	1.8	188	1	COAT_CAYMV	P19128 cacao yello
C 99	7	1.8	188	1	HXD8_CHICK	P23459 gallus gall
C 100	7	1.8	193	1	HUNB_DROIK	046242 drosophila
C 101	7	1.8	194	1	PRAN_HELAU	018641 helicoverpa
C 102	7	1.8	194	1	PRAN_HELZE	P11159 heliothis z
C 103	7	1.8	195	1	GRP3_DAUCA	P37705 daucus caro
C 104	7	1.8	195	1	IRX6_HUMAN	P78412 homo sapien
C 105	7	1.8	195	1	KITH_BACNA	052951 bacillus su
C 106	7	1.8	198	1	YF91_MYCLE	Q49626 mycobacteri
C 107	7	1.8	201	1	COX2_BRUME	Q8yfg7 brucella me
C 108	7	1.8	201	1	COX2_BRUSU	Q8g261 brucella su
C 109	7	1.8	204	1	AR11_XENLA	Q31828 xenopus lae
C 110	7	1.8	209	1	VAD2_TREPA	083539 treponema p
C 111	7	1.8	212	1	RL14_HUMAN	P50914 homo sapien
C 112	7	1.8	214	1	CMB2_DIACA	Q42498 dianthus ca
C 113	7	1.8	215	1	HAN1_RABIT	P57100 oryctolagus
C 114	7	1.8	215	1	PDX5_CERAE	Q9glw7 cercopithec
C 115	7	1.8	215	1	PDX5_PAPHA	Q9glw9 papio hamad
C 116	7	1.8	220	1	ATPO_NEUCR	Q9p602 neurospora
C 117	7	1.8	222	1	TPIS_METAC	Q8thb0 methanosarc
C 118	7	1.8	222	1	TPIS_METMA	Q8pxe2 methanosarc
C 119	7	1.8	223	1	TPIS_ARCFU	Q28965 archaeoglob
C 120	7	1.8	224	1	NORD_PSAE	Q9h2k9 pseudomonas
C 121	7	1.8	226	1	DLIP_MOUSE	Q9dca7 mus musculu
C 122	7	1.8	226	1	DLIP_RAT	P58821 rattus norv
C 123	7	1.8	227	1	AGI_ORYSA	P11219 oryza sativ
C 124	7	1.8	229	1	TPIS_METH	Q27120 methanobact
C 125	7	1.8	231	1	ASCI_MOUSE	Q20267 mus musculu
C 126	7	1.8	233	1	ASCI_RAT	P19359 rattus norv
C 127	7	1.8	236	1	ASCI_HUMAN	P50553 homo sapien
C 128	7	1.8	239	1	US21_HCNVA	P09723 human cytom
C 129	7	1.8	240	1	RSMN_HUMAN	P14648 homo sapien
C 130	7	1.8	243	1	AX28_SOYBN	P13089 glycine max
C 131	7	1.8	243	1	RS3_SYNPE	Q24695 synchococ
C 132	7	1.8	249	1	VITM_MANSE	P19616 manduca sex
C 133	7	1.8	253	1	TPIS_BORBU	Q59182 borrelia bu
C 134	7	1.8	260	1	RS3_ANASP	Q8yp15 anabaena sp
C 135	7	1.8	261	1	COX3_MACRO	P92665 macropus ro

RESULT 1

NUP62\_MOUSE

NU62\_MOUSE

STANDARD;

PR7; 526 AA.

Q63850; Q99JN7;

28-FEB-2003 (Rel. 41, Created)

28-FEB-2003 (Rel. 41, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)

Nuclear pore glycoprotein p62 (62 kDa nucleoporin).

NUP62.

Mus musculus (Mouse).

OS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxId=10090;

[1]

SEQUENCE FROM N.A.

MEDLINE=92007945; PubMed=1915419;

Cordes V., Maizenege I., Krohne G.;

"Nuclear pore complex glycoprotein p62 of xenopus laevis and mouse:

cDNA cloning and identification of its glycosylated region.";

Eur. J. Cell Biol. 55:31-47(1991).

[2]

SEQUENCE FROM N.A.

MEDLINE=23388257; PubMed=12477932;

Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

```
RESULT 2
PROS_DROVI
ID PROS_DROVI STANDARD; PRT; 1556 AA.
AC Q9U6A1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein prospero.
GN PROS.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7244;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20503846; PubMed=11051550;
RA Xu C., Kauffmann R.C., Zhang J., Kladny S., Carthew R.W.;
RT "Overlapping activators and repressors delimit transcriptional
RT response to receptor tyrosine kinase signals in the Drosophila eye.";
RL Cell 103:87-97(2000).
CC -!- FUNCTION: Required for proper neuronal differentiation of most or
CC all neurons and their precursors in central and peripheral nervous
CC systems, axonal outgrowth and pathfinding. Not required for the
CC specification of neuronal identity. May regulate transcription by
CC binding to DNA (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity)
CC -!- SIMILARITY: Belongs to the Prospero homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF190405; AAF06660.1; -
CC FlyBase; FBgn028753; Dvir\pros.
CC InterPro; IPR007738; Prox1.
CC Pfam; PF05044; Prox1; 1.
CC
CC Nuclear protein; Transcription regulation; DNA-binding;
CC Homeobox; Developmental protein.
CC
CC FT DOMAIN 4 12 POLY-ALA.
CC FT DOMAIN 19 36 ASN-RICH.
CC FT DOMAIN 150 172 ASN-RICH.
CC FT DOMAIN 206 209 POLY-ALA.
CC FT DOMAIN 237 292 GLN-RICH.
CC FT DOMAIN 305 309 POLY-ASN.
CC FT DOMAIN 349 381 SER-RICH.
CC FT DOMAIN 443 473 ASP-RICH.
CC FT DOMAIN 544 547 POLY-ALA.
CC FT DOMAIN 574 1080 GLN-RICH.
CC FT DOMAIN 888 906 ASN-RICH.
CC FT DOMAIN 1027 1030 POLY-ALA.
CC FT DOMAIN 1045 1054 POLY-GLN.
CC FT DOMAIN 1057 1062 POLY-GLN.
CC FT DOMAIN 1132 1189 THR-RICH.
CC FT DOMAIN 1140 1145 POLY-ALA.
CC FT DOMAIN 1154 1163 POLY-GLN.
CC FT DOMAIN 1183 1189 POLY-THR.
CC FT DOMAIN 1090 1097 NUCLEAR LOCALIZATION SIGNAL (BY
CC SIMILARITY).
CC FT DOMAIN 1330 1337 POLY-GLN.
CC FT DNA_BIND 1394 1456 HOMEBOX (ATYPICAL) (BY SIMILARITY).
CC FT DOMAIN 1457 1556 PROSPERO-LIKE (BY SIMILARITY).
CC
CC SEQUENCE 1556 AA; 171029 MW; 6FEACFEA2D73E644 CRC64;
SQ
Alignment Scores:
Pred. No.: 0.299 Length: 1556
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
```

```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.63% Indels: 0
DB: 1 Gaps: 0
US-09-729-264-1 (1-1175) x PROS_DROVI (1-1556)
QY 836 ACACACGACGCGCGCAGCAGCAGCAGCG 807
DB 1136 ThrThrThrAlaAlaAlaAlaAla 1145
RESULT 3
TYRT_STRAL
ID TYRT_STRAL STANDARD; PRT; 126 AA.
AC P5046;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Tyrosinase co-factor.
GN MELC1.
OS Streptomyces albus G.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1962;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 40480;
RA Wehmeier U.F., Brass N., Roessler C., Piepersberg W.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN MAY FUNCTION TO DELIVER COPPER TO
CC TYROSINASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X95705; CAA65004.1; -
CC Melanin biosynthesis; Copper.
CC
CC SEQUENCE 126 AA; 12916 MW; 1785CC2C777C0106 CRC64;
SQ
Alignment Scores:
Pred. No.: 4.13 Length: 126
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.37% Indels: 0
DB: 1 Gaps: 0
US-09-729-264-1 (1-1175) x TYRT_STRAL (1-126)
QY 860 ACAACGGCAGCAGCAGTTCAGCCACA 834
DB 11 ThrThrAlaAlaAlaValAlaAlaThr 19
RESULT 4
PLAS_SILPR
ID PLAS_SILPR STANDARD; PRT; 165 AA.
AC P07030;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Plastocyanin, chloroplast precursor.
GN PETE.
OS Silene pratensis (White campion) (lychnis alba).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Silene.
OX NCBI_TaxID=52853;
RN [1]
RP SEQUENCE FROM N.A.
```

RA Sneekens S., de Groot M., van Binsbergen J., Weisbeek P.J.;  
 RT "Sequence of the precursor of the chloroplast thylakoid lumen protein  
 RT plastocyanin.";  
 RL Nature 317:456-458(1985).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).  
 RX MEDLINE=92238710; PubMed=10220581;  
 RA Sugawara H., Inoue T., Li C., Gotowda M., Hibino T., Takabe T.,  
 RA Kai Y.;  
 RT "Crystal structures of wild-type and mutant plastocyanins from a  
 RT higher plant, *Silene*.";  
 RL J. Biochem. 125:899-903(1999).  
 CC -1- FUNCTION: Participates in electron transfer between P700 and the  
 CC cytochrome b6-f complex in photosystem I.  
 CC -1- SUBCELLULAR LOCATION: Loosely bound to the inner thylakoid  
 CC membrane surface in chloroplasts.  
 CC -1- SIMILARITY: Contains 1 plastocyanin-like domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X02965; CAA26709.1; -;  
 DR PIR; A24404; CUQH.  
 DR PDB; 1BYO; 15-OCT-99.  
 DR PDB; 1BYX; 19-OCT-99.  
 DR InterPro; IPR000923; BlueCu\_1.  
 DR InterPro; IPR001235; Copper\_Blue.  
 DR InterPro; IPR008972; Cupredoxin.  
 DR Pfam; PF00127; copper-binding; 1.  
 DR PRINTS; PR00156; COPPERBLUE.  
 DR ProDom; PD001235; Copper\_blue; 1.  
 DR PROSITE; PS00196; COPPER\_BLUE; 1.  
 DR Chloroplast; Electron transport; Copper; Thylakoid; Membrane;  
 KW Transit peptide; 3D-structure.  
 FT TRANSIT 1 66 CHLOROPLAST.  
 FT CHAIN 67 165 PLASTOCYANIN.  
 FT DOMAIN 67 165 PLASTOCYANIN-LIKE.  
 FT METAL 103 103 COPPER (BY SIMILARITY).  
 FT METAL 150 150 COPPER (BY SIMILARITY).  
 FT METAL 153 153 COPPER (BY SIMILARITY).  
 FT METAL 158 158 COPPER (BY SIMILARITY).  
 SQ SEQUENCE 165 AA; 16620 MW; C4F817E69BC514A0 CRC64;  
 Alignment Scores:  
 Pred. No.: 3.97 Length: 165  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.31% Indels: 0  
 DB: 1 Gaps: 0  
 US-09-729-264-1 (1-1175) x PLAS\_SILPR (1-165)  
 QY 829 GTTGTCTGGCTGCAACTGCTGCTGCC 855  
 Db 49 ValValValAlaAlaThrAlaAlaAla 57  
 RESULT 5  
 ID TGAS STRMB STANDARD; PRT; 407 AA.  
 AC P81453; Q8KRJ2; Q9ZAF5;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DE Protein-glutamine gamma-glutamyltransferase precursor (EC 2.3.2.13)  
 DE (transglutaminase) (TGase) (MTG).  
 OS Streptomyces mobaraensis (Streptovorticillium mobaraense).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycineae; Streptomycetaceae; Streptomycetes.  
 OX NCBI\_TaxID=35621;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-IFO 13819;  
 RX MEDLINE=22401437; PubMed=12514016;  
 RA Kikuchi Y., Date M., Yokoyama K.-I., Umezawa Y., Matsui H.;  
 RT "Secretion of active-form Streptovorticillium mobaraense  
 RT pro-transglutaminase by Corynebacterium glutamicum: processing of the  
 RT Streptomycetes albogriseolus.";  
 RL Appl. Environ. Microbiol. 69:358-366(2003).  
 RN [2]  
 RP SEQUENCE OF 32-407 FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=DSMZ 40587;  
 RX MEDLINE=99053680; PubMed=9839945;  
 RA Pasternack R., Dorsch S., Otterbach J.T., Robenek I.R., Wolf S.,  
 RA Fuchsbaue H.-L.;  
 RT "Bacterial pro-transglutaminase from Streptovorticillium mobaraense:  
 RT purification, characterisation and sequence of the zymogen.";  
 RL Eur. J. Biochem. 257:570-576(1998).  
 RN [3]  
 RP SEQUENCE OF 77-407, AND MASS SPECTROMETRY.  
 RC STRAIN-S-8112;  
 RX MEDLINE=93280110; PubMed=8099353;  
 RA Kanaji T., Ozaki H., Takao T., Kawajiri H., Ide H., Motoki M.,  
 RA Shimomishi Y.;  
 RT "Primary structure of microbial transglutaminase from  
 RT Streptovorticillium sp. strain s-8112.";  
 RL J. Biol. Chem. 268:11565-11572(1993).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 77-407.  
 RC STRAIN=IFO 13819;  
 RX MEDLINE=22313549; PubMed=12221081;  
 RA Kashiwagi T., Yokoyama K.-I., Iehikawa K., Ono K., Ejima D.,  
 RA Matsui H., Suzuki E.;  
 RT "Crystal structure of microbial transglutaminase from  
 RT Streptovorticillium mobaraense.";  
 RL J. Biol. Chem. 277:44252-44260(2002).  
 CC -1- FUNCTION: Catalyzes the cross-linking of proteins and the  
 CC conjugation of polyamines to proteins.  
 CC -1- CATALYTIC ACTIVITY: Protein glutamine + alkylamine = protein N(S)-  
 CC alkylglutamine + NH(3).  
 CC -1- MASS SPECTROMETRY: MW=37869.2; MW\_ERR=8.8; METHOD=Electrospray;  
 CC RANGE=77-407.  
 CC -1- BIOTECHNOLOGY: Sold under the name Activa TG by Ajinomoto. It has  
 CC the ability to crosslink protein molecules present in food without  
 CC the use of salt or binders. Used to improve some of the physical  
 CC properties such as firmness, elasticity and moisture retention of  
 CC food such as meat, poultry and seafood.  
 CC -1- SIMILARITY: Belongs to the bacterial TGase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AF531437; AAM95951.1; -;  
 DR EMBL; Y18315; CAA77128.1; -;  
 DR PDB; 1IU4; 04-MAR-03.  
 KW Transglutaminase; Acyltransferase; Zymogen; 3D-structure; Signal.  
 FT SIGNAL 1 31 POTENTIAL.  
 FT PROPEP 32 76  
 FT CHAIN 77 407 PROTEIN-GLUTAMINE GAMMA-  
 FT GLUTAMYLTRANSFERASE.  
 FT ACT SITE 140 140  
 FT ACT SITE 331 331  
 FT ACT SITE 350 350  
 SQ SEQUENCE 407 AA; 45684 MW; 10F7F7A04EAB2DF4 CRC64;



Alignment Scores: 3.46 Length: 407  
Pred. No.: 9.00 Matches: 9  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 2.31% Gaps: 0  
DB: 1

US-09-729-264-1 (1-1175) x TCAS\_STRMB (1-407)  
QY 356 CQTRACCTCAAGTTATGGAGACT 382  
Db 95 ProTyArgProSerTyrGlyArgala 103

RESULT 6  
MUC1\_XENLA  
ID MUC1\_XENLA STANDARD; PRT; 662 AA.  
AC Q05049;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Integumentary mucin C.1 (FIM-C.1) (Fragment).  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4; 5; 6 AND 7).  
RC TISSUE=Skin;  
RX MEDLINE=93077556; PubMed=1447205;  
RA Hauser F., Hoffmann W.;  
RT "P-domains as shuffled cysteine-rich modules in integumentary mucin C.1 (FIM-C.1) from Xenopus laevis. Polydispersity and genetic polymorphism.";  
RL J. Biol. Chem. 267:24620-24624 (1992).  
CC -!- FUNCTION: Could be involved in defense against microbial infections. Protects the epithelia from external environment.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=7;  
CC Comment=Additional isoforms seem to exist. Experimental confirmation may be lacking for some isoforms;  
CC Name=1;  
CC IsoId=Q05049-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q05049-2; Sequence=VSP\_004650;  
CC Name=3;  
CC IsoId=Q05049-3; Sequence=VSP\_004651;  
CC Name=4;  
CC IsoId=Q05049-4; Sequence=VSP\_004647, VSP\_004648;  
CC Name=5;  
CC IsoId=Q05049-5; Sequence=VSP\_004646, VSP\_004649, VSP\_004650;  
CC Name=6;  
CC IsoId=Q05049-6; Sequence=VSP\_004646, VSP\_004648;  
CC Name=7;  
CC IsoId=Q05049-7; Sequence=VSP\_004647;  
CC -!- TISSUE SPECIFICITY: Skin.  
CC -!- PTM: Extensively O-glycosylated.  
CC -!- SIMILARITY: Contains 6 p-type (trefoil) domains.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
CC EMBL; L02115; AAA74725.1; -.  
DR PIR; A45155; A45155.  
DR HSP; P01359; 2PSP.  
DR InterPro; IPR000519; P\_trefoil.

DR PFam; PF00088; trefoil; 6.  
DR PRINTS; PRO0680; PTREFOIL.  
DR SMART; SM00018; PD; 6.  
DR PROSITE; PS00025; P\_TREFOIL; 6.  
KW Repeat; Glycoprotein; Alternative splicing.  
FT NON TER 1  
FT DOMAIN 81 144 8 X 8 AA APPROXIMATE TANDEM REPEATS, ALA/THR-RICH.  
FT REPEAT 81 88 1-1.  
FT REPEAT 89 96 1-2.  
FT REPEAT 97 104 1-3.  
FT REPEAT 105 112 1-4.  
FT REPEAT 113 120 1-5.  
FT REPEAT 121 128 1-6.  
FT REPEAT 129 136 1-7.  
FT REPEAT 137 144 1-8.  
FT DOMAIN 161 202 P-TYPE 1.  
FT DOMAIN 218 301 8 X APPROXIMATE TANDEM REPEATS, THR-RICH.  
FT REPEAT 218 224 2-1.  
FT REPEAT 225 239 2-2.  
FT REPEAT 240 249 2-3.  
FT REPEAT 250 259 2-4.  
FT REPEAT 260 275 2-5.  
FT REPEAT 276 287 2-6.  
FT REPEAT 288 294 2-7.  
FT REPEAT 295 301 2-8.  
FT DOMAIN 306 347 P-TYPE 2.  
FT DOMAIN 353 394 P-TYPE 3.  
FT DOMAIN 402 522 12 X APPROXIMATE TANDEM REPEATS, THR-RICH.  
FT REPEAT 402 411 3-1.  
FT REPEAT 412 419 3-2.  
FT REPEAT 420 431 3-3.  
FT REPEAT 432 443 3-4.  
FT REPEAT 444 453 3-5.  
FT REPEAT 454 460 3-6.  
FT REPEAT 461 472 3-7.  
FT REPEAT 473 479 3-8.  
FT REPEAT 480 491 3-9.  
FT REPEAT 492 498 3-10.  
FT REPEAT 499 515 3-11.  
FT REPEAT 516 522 3-12.  
FT DOMAIN 525 566 P-TYPE 4.  
FT DOMAIN 572 613 P-TYPE 5.  
FT DOMAIN 620 661 P-TYPE 6.  
FT DISULFID 162 188 BY SIMILARITY.  
FT DISULFID 172 187 BY SIMILARITY.  
FT DISULFID 182 199 BY SIMILARITY.  
FT DISULFID 307 333 BY SIMILARITY.  
FT DISULFID 317 332 BY SIMILARITY.  
FT DISULFID 327 344 BY SIMILARITY.  
FT DISULFID 354 380 BY SIMILARITY.  
FT DISULFID 364 379 BY SIMILARITY.  
FT DISULFID 374 391 BY SIMILARITY.  
FT DISULFID 526 552 BY SIMILARITY.  
FT DISULFID 536 551 BY SIMILARITY.  
FT DISULFID 546 563 BY SIMILARITY.  
FT DISULFID 573 599 BY SIMILARITY.  
FT DISULFID 583 598 BY SIMILARITY.  
FT DISULFID 593 610 BY SIMILARITY.  
FT DISULFID 621 647 BY SIMILARITY.  
FT DISULFID 631 646 BY SIMILARITY.  
FT DISULFID 641 658 Missing (in isoform 5 and isoform 6).  
FT VARSPLIC 240 259 /FTid=VSP\_004646.  
FT VARSPLIC 250 259 Missing (in isoform 4 and isoform 7).  
FT VARSPLIC 276 294 /FTid=VSP\_004647.  
FT VARSPLIC 278 278 Missing (in isoform 4 and isoform 6).  
FT VARSPLIC 278 278 Missing (in isoform 5).  
FT VARSPLIC 306 350 /FTid=VSP\_004649.  
FT VARSPLIC 306 350 Missing (in isoform 2 and isoform 5).  
FT VARSPLIC 306 350 /FTid=VSP\_004650.

FT VARSPLIC 420 498 Missing (in isoform 3).  
FT FTID=VSP\_004651.  
FT VARIANT 276 276 K -> E.  
FT VARIANT 354 354 C -> R.  
FT VARIANT 415 415 T -> A.  
SQ SEQUENCE 662 AA; 67774 MW; F085277F1ED2FD40 CRC64;

Alignment Scores:  
Pred. No.: 3.21 Length: 662  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.37% Indels: 0  
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x MUCI\_XENLA (1-662)

QY 860 ACAAGCGCAGCAGTTCAGGCACA 834  
Db 3 ThrAlaAlaAlaValAlaAlaThr 11

## RESULT 7

FLBA\_EMENI STANDARD; PRT; 719 AA.

ID FLBA\_EMENI

AC P38093;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Developmental regulator fibA.

GN FLBA.

OS Emericella nidulans (Aspergillus nidulans).

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eukaryotes; Trichocomaceae; Emericella.

OX NCBI\_TaxID=162425;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95131754; PubMed=7830576;

RA Lee B.N., Adams T.H.;

RT "Overexpression of fibA, an early regulator of Aspergillus asexual sporulation, leads to activation of brIA and premature initiation of development";

RL Mol. Microbiol. 14:323-334(1994).

CC -1- FUNCTION: Required for asexual sporulation and normal colony development. May be involved in brIA activation. Could play a regulatory role in controlling the flag-initiated signal transduction pathway that triggers the asexual reproduction.

CC -1- DEVELOPMENTAL STAGE: Present throughout the asexual cycle.

CC -1- SIMILARITY: Contains 1 RGS domain.

CC -1- SIMILARITY: Contains 1 DEP domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: L24395; AAT73955.1; -.

PIR: S60771; S60771.

HSP: P49795; ICMZ.

InterPro: IPR000591; DEP.

InterPro: IPR000342; Regl\_Gprotein.

Pfam: PF00610; DEP; 1.

Pfam: PF00615; RGS; 1.

PRINTS: PR01301; RGS-PROTEIN.

SMART: SM00049; DEP; 2.

SMART: SM00315; RGS; 1.

PROSITE: PS0186; DEP; 1.

PROSITE: PS0132; RGS; 1.

Signal transduction inhibitor.

DOMAIN 38 44

FT POLY-ALA.

FT POLY-THR.

FT DOMAIN 425 511 DEP.  
FT DOMAIN 540 685 RGS.  
SQ SEQUENCE 719 AA; 78798 MW; 7AB5D3ABC0B53AC8 CRC64;

Alignment Scores:  
Pred. No.: 3.17 Length: 719  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x FLBA\_EMENI (1-719)

QY 808 GCTGCTGCTGCTGCCGCGCTGTTGT 834

Db 39 AlaAlaAlaAlaAlaAlaValVal 47

## RESULT 8

BRH2 HUMAN

ID BRH2 HUMAN

AC Q9NY43;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE BarH-like 2 homeobox protein (Fragment).

GN BARHL2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20275633; PubMed=10814725;

RA Bulfone A., Menguzzato E., Broccoli V., Marchitelli A., Gattuso C.,

Mariani M., Consalez G.G., Martinez S., Ballabio A., Banfi S.;

RT "Barhl1, a gene belonging to a new subfamily of mammalian homeobox

genes, is expressed in migrating neurons of the CNS.";

Hum. Mol. Genet. 9:1443-1452(2000).

RL Hum. Mol. Genet. 9:1443-1452(2000).

CC -1- FUNCTION: Potential regulator of neural basic helix-loop-helix

genes (by similarity).

CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -1- SIMILARITY: Belongs to the BAR homeobox family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AJ251753; CAB92440.1; -.

TRANSFAC: T03923; -.

DR GENE: HGNC:954; BARHL2.

DR MIM: 605212; -.

DR InterPro: IPR001356; Homeobox.

DR PROSITE: PS00027; HOMEBOX 1; PARTIAL.

DR PROSITE: PS00071; HOMEBOX 2; PARTIAL.

DR Homeobox; Transcription regulation; DNA-binding; Nuclear protein.

FT NON\_TER 1

FT NON\_TER 68

SQ SEQUENCE 68 AA; 7388 MW; 597AE034911DF997 CRC64;

Alignment Scores:

Pred. No.: 42.8 Length: 68

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.11% Indels: 0

DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x BRH2\_HUMAN (1-68)

```
QY 830 ACGACGGCGGACGACGACGACG 807
|||||
Db 32 ThrThrAlaAlaAlaAlaAla 39

RESULT 9
RLA2_LEIDO
ID_RLA2_LEIDO STANDARD; PRT; 105 AA.
AC 043940;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60S acidic ribosomal protein P2 (Acidic ribosomal protein-1).
GN ARP-1.
OS Leishmania donovani.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5661;
RN STRAIN=IS Sudanese;
RC SEQUENCE FROM N.A.
RA Cheng J., Zhao W., Melby P.C.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Plays an important role in the elongation step of
protein synthesis.
CC -!- SUBUNIT: P1 and P2 exist as dimers at the large ribosomal subunit.
CC -!- PTM: Phosphorylated (By similarity).
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF034539; AAB88451.1; -
DR InterPro; IPR001813; Ribosomal 60S.
DR InterPro; IPR001859; Ribosomal P2.
DR Pfam; PF00428; 60S ribosomal; I.
DR PRINTS; PR00456; RIBOSOMALP2.
KW Ribosomal protein; Phosphorylation.
SQ SEQUENCE 105 AA; 10446 MW; 181CD854C34345BE6 CRC64;

Alignment Scores:
Pred. No.: 40.1 Length: 105
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.06% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x RLA2_LEIDO (1-105)

QY 808 GCTGCTGCTGCGCGCGCTGCTT 831
|||||
Db 76 AlaAlaAlaAlaAlaAlaVal 83

RESULT 10
SSB_DROME
ID_SSB_DROME STANDARD; PRT; 146 AA.
AC P54622; Q9V3U4;
DT 01-OCT-1996 (Rel. 34, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Single-stranded DNA-binding protein, mitochondrial precursor (Mt-SSB)
DE (MtSSB) (Dm mt-SSB) (low power protein).
GN MTSB OR LOPO OR CG4337.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
```

```
RN SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RC STRAIN=Oregon-R;
RX MEDLINE=20250922; PubMed=10788480;
RA Ruiz de Mena I., Lefai E., Garesse R., Kaguni L.S.;
RT "Regulation of mitochondrial single-stranded DNA-binding protein gene
expression links nuclear and mitochondrial DNA replication in
Drosophila.";
RL J. Biol. Chem. 275:13628-13636(2000).

RN [2]
SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).

RN [3]
REVIEWS.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Beriman B.P.,
Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).

RN [4]
SEQUENCE OF 4-146 FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND
DEVELOPMENTAL STAGE.
RC STRAIN=Canton-S; TISSUE=Ovary;
RX MEDLINE=94266149; PubMed=8206370;
RA Stroumbakis N.D., Li Z., Tolias P.P.;
RT "RNA- and single-stranded DNA-binding (SSB) proteins expressed during
Drosophila melanogaster oogenesis: a homolog of bacterial and
eukaryotic mitochondrial SSBs.";
```

Gene 143:171-177(1994).

[5]  
 RL SEQUENCE OF 23-51, FUNCTION, SUBUNIT, SUBCELLULAR LOCATION, AND MASS  
 RP SPECTROMETRY.  
 RX MEDLINE=95403346; PubMed=7673145;  
 RA Thomas P., Farr C.L., Marton R.F., Kaguni L.S., Cotterill S.;  
 RT "Mitochondrial single-stranded DNA-binding protein from *Drosophila*  
 embryos. Physical and biochemical characterization.";  
 RL J. Biol. Chem. 270:21137-21143(1995).

[6]  
 RL FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
 RP MEDLINE=21191852; PubMed=11294869;  
 RA Maier D., Farr C.L., Poock B., Alahari A., Vogel M., Fischer S.,  
 Kaguni L.S., Schneuwly S.;  
 RT "Mitochondrial single-stranded DNA-binding protein is required for  
 mitochondrial DNA replication and development in *Drosophila*  
 melanogaster.";  
 RL Mol. Biol. Cell 12:821-830(2001).

CC -1- FUNCTION: This protein binds preferentially and cooperatively to  
 CC pyrimidine rich ss-DNA. Required for mitochondrial DNA  
 CC replication.  
 CC -1- SUBUNIT: Homotetramer.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -1- TISSUE SPECIFICITY: Uniformly distributed in the early embryo.  
 CC High levels detected in the anterior and posterior midgut  
 CC primordia of stage 12 embryos. In larvae, high levels were  
 CC detected in proliferating tissues including the CNS and digestive  
 CC tract. In adults, highly expressed in the CNS, digestive tract and  
 CC ovary.  
 CC -1- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.  
 CC Levels are high during embryogenesis and in the larvae but  
 CC decrease in the pupae before increasing again in the adult.  
 CC -1- MASS SPECTROMETRY: MW=13845; MW\_ERR=14; METHOD=MALDI;  
 CC RANGE=23-146.  
 CC -1- SIMILARITY: Contains 1 SSB domain.  
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to erroneous  
 CC gene model prediction.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; AF181084; AAF16936.1; ALT\_SEQ.  
 DR EMBL; AE003712; AAF55287.2; -.  
 DR EMBL; U00669; AAA20507.1; ALT\_INIT.  
 DR HSPSP; Q04837; 3ULL.  
 DR GO; GO:0000262; C:mitochondrial chromosome; IDA.  
 DR GO; GO:0005739; C:mitochondrion; IDA.  
 DR GO; GO:0003697; F:single-stranded DNA binding; IDA.  
 DR GO; GO:0006264; P:mitochondrial DNA replication; IDA.  
 DR GO; GO:0000002; P:mitochondrial genome maintenance; IDA.  
 DR InterPro; IPR000424; SSB\_protein.  
 DR Pfam; PF00436; SSB; 1.  
 DR TIGRfam; TIGR00621; ssb; 1.  
 DR PROSITE; PS50935; SSB; 1.  
 KW DNA-binding; DNA replication; Mitochondrion; Transit peptide.  
 FT TRANSIT 1 22 MITOCHONDRION.  
 FT CHAIN 23 146 SINGLE-STRANDED DNA-BINDING PROTEIN.  
 FT DOMAIN 38 142 SSB.  
 SQ SEQUENCE 146 AA; 16367 MW; AD505175C0555D48 CRC64;

Alignment Scores:  
 Pred. No.: 38.1 Length: 146  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.11% Indels: 0  
 DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x SSB\_DROME (1-146)

QY 839 GCCACACACGACGGCGGCGAGCA 816  
 Db 23 AlathrThrThrAlaAlaAla 30  
 |||||

RESULT 11  
 HUNB\_DROMM STANDARD; PRT; 158 AA.  
 ID HUNB\_DROMM STANDARD; PRT; 158 AA.  
 AC 046248; 046249;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hunchback protein (Fragments).  
 GN HB.  
 OS *Drosophila mimica* (Fruit fly) (*Idiomyia mimica*).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 CX NCBI\_TaxID=7270;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Baker R.H., Desalle R.;  
 RT "Multiple sources of character information and the phylogeny of  
 RT Hawaiian *Drosophilids*.";  
 RL Syst. Biol. 46:654-673(1997).  
 CC -1- FUNCTION: Gap class segmentation protein that controls development  
 CC of head structures (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; U93012; AAC03260.1; -.  
 DR EMBL; U93013; AAC03261.1; -.  
 DR FlyBase; FBgn003764; Dmichb.  
 KW Developmental protein; Gap protein; Zinc-finger;  
 KW Metal-binding; DNA-binding; Repeat; Nuclear protein.  
 FT NON TER 1 1  
 FT DOMAIN 17 31 POLY-HIS.  
 FT DOMAIN 60 64 POLY-GLN.  
 FT NON CONS 64 65  
 FT DOMAIN 91 95  
 FT DOMAIN 96 99  
 FT NON TER 158 158  
 SQ SEQUENCE 158 AA; 17576 MW; 114B650BD4DC8CDE CRC64;

Alignment Scores:  
 Pred. No.: 37.7 Length: 158  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.11% Indels: 0  
 DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x HUNB\_DROMM (1-158)

QY 836 ACACACACGACGGCGGCGAGCA 813  
 Db 92 ThrThrThrThrAlaAlaAla 99  
 |||||

RESULT 12  
 YC17\_DROME STANDARD; PRT; 184 AA.  
 ID YC17\_DROME STANDARD; PRT; 184 AA.  
 AC Q9VH95;

DR	FlyBase; FBgm0037728; CG16817.
KW	InterPro; IPR009878; HSP20_Chap.
DW	Hypothetical protein.
FT	DOMAIN 145 184
SQ	SEQUENCE 184 AA; 20753 MW; 982909B6E25SDB08 CRC64;
 Alignment Scores:	
Pred. No.:	36.8 Length: 184
Score:	8.00 Matches: 8
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	2.11% Indels: 0
DB:	1 Gaps: 0
 US-09-729-264-1 (1-1175) x YC17_DROME (1-184)	
QY	688 TTAATACCACTCCAGTGTCTTGG 665
Db	7 LeulleProProValSerTip 14
 RESULT 13	
MMS3 MYCTU	
ID _MMS3 MYCTU STANDARD; PRT; 299 AA.	
IC Q10330;	
DT 01-OCT-1996 (Rel. 34, Created)	
DT 01-OCT-1996 (Rel. 34, Last sequence update)	
DT 10-OCT-2003. (Rel. 42, Last annotation update)	
DE Putative membrane protein MWP53	
GN MWP53 OR RV2198C OR MT2254 OR MTCV190.09C OR MB2221C.	
OS Mycobacterium tuberculosis, and	
OS Mycobacterium bovis.	
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX NCBI_TaxID=1773, 1765;	
RN [1]	
RP SPECIES FROM N.A.	
RC SPECIES=M.tuberculosis; STRAIN=H37RV;	
RA MEDLINE=98295987; PubMed=9634230;	
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,	
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,	
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,	
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,	
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,	
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,	
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,	
RA Sulston J.E., Taylor K., Whitehead S., Barrall E.G.;	
RT "Deciphering the biology of Mycobacterium tuberculosis from the	
RL complete genome sequence."	
RL Nature 393:537-544(1998).	
[2]	
RP SEQUENCE FROM N.A.	
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;	
RA MEDLINE=22206494; PubMed=12218036;	
RX Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,	
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,	
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,	
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,	
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;	
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and	
RL laboratory strains."	
RL J. Bacteriol. 184:5479-5490(2002).	
[3]	
RP SEQUENCE FROM N.A.	
RC SPECIES=M.bovis; STRAIN=AF2122/97;	
RA MEDLINE=22709107; PubMed=12788972;	
RX Garner T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,	
RA Pryor M., Duthoy S., Gronlund S., Lacroix C., Monsemp C., Simon S.,	
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,	
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;	
RT "the complete genome sequence of Mycobacterium bovis."	
RL Proc. Natl. Acad. Sci. U.S.A. 100:7897-7882(2003).	
CC -!- SIMILARITY: Belongs to the mmps family.	

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; Z70283; CAA94267.1; --  
 CC DR EMBL; AE007071; AAK46540.1; --  
 CC DR EMBL; BX248341; CAD97074.1; --  
 CC DR PIR; G70784; G70784.  
 CC TIGR; MT2254; --

CC Hypothetical protein; Transmembrane; Complete proteome.  
 CC KW Hypothetical protein; Transmembrane; Complete proteome.  
 CC FT TRANSMEM 101 121 POTENTIAL.  
 CC FT DOMAIN 146 221 PRO/THR-RICH.  
 CC SQ SEQUENCE 299 AA; 30955 MW; DBF65DD4D2E8FBD7 CRC64;

Alignment Scores:  
 Pred. No.: 34.2 Length: 299  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.11% Indels: 0  
 DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x MMS3\_MYCTU (1-299)

QY 839 GCCACACACGCGCGGCGAGCA 816

DB 193 AlathrThrThrAlaAlaAla 200

RESULT 14

ALB3\_MAIZE  
 ID\_ALB3\_MAIZE STANDARD; PRT; 303 AA.  
 AC P10593;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Albumin b-32 protein (EC 3.2.2.22) (Opaque-6 protein) (rRNA N-glycosidase).  
 DE O6.  
 GN Zea mays (Maize).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 OX NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91346687; PubMed=2102870;  
 RA Hartings H., Lazaroni N., Marsan P.A., Aragay A., Thompson R.,  
 RA Salamini F., di Fonzo N., Palau J., Motto M.,  
 RT "The b-32 protein from maize endosperm: characterization of genomic  
 RT sequences encoding two alternative central domains.";  
 RL Plant Mol. Biol. 14:1031-1040(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88334499; PubMed=3419419;  
 RA di Fonzo N., Hartings H., Brembilla M., Motto M., Soave C.,  
 RA Navarro E., Palau J., Rhode W., Salamini F.,  
 RT "The b-32 protein from maize endosperm, an albumin regulated by the  
 RT O2 locus: nucleic acid (cDNA) and amino acid sequences.";  
 RL Mol. Gen. 212:461-487(1988).  
 CC -1- FUNCTION: A possible regulatory factor for the synthesis of zeins,  
 CC the major group of storage proteins.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of the N-glycosidic bond at one  
 CC specific adenosine on the 28S rRNA.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: Endosperm.  
 CC -1- SIMILARITY: Belongs to the ribosome-inactivating protein family.

Type 1 RIP subfamily.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; X54212; CAA38124.1; --  
 CC DR EMBL; X07987; CAA30797.1; --  
 CC DR PIR; S03172; S03172.  
 CC DR MaizeDB; 30000; --  
 CC DR InterPro; IPR001574; RIP.  
 CC Pfam; PF00161; RIP; 1  
 CC PRINTS; PR00396; SHIGARICIN.  
 CC DR PROSITE; PS00275; SHIGA RICIN; 1.

CC KW Plant defense; Protein synthesis inhibitor; Hydrolase; Toxin.  
 CC FT DOMAIN 182 188 POLY-ALA.  
 CC FT DOMAIN 286 294 POLY-ALA.  
 CC FT CONFLICT 269 269 D -> N (IN REF. 2).  
 CC SQ SEQUENCE 303 AA; 32428 MW; 24003521CEFP91790 CRC64;

Alignment Scores:  
 Pred. No.: 34.1 Length: 303  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.11% Indels: 0  
 DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x ALB3\_MAIZE (1-303)

QY 830 ACGACGCGCGACGACGACGCG 807

DB 284 ThrThrAlaAlaAlaAlaAla 291

RESULT 15

RIP9\_MAIZE  
 ID\_RIP9\_MAIZE STANDARD; PRT; 304 AA.  
 AC P25892;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Ribosome-inactivating protein 9 (EC 3.2.2.22) (rRNA N-glycosidase)  
 DE (B-32 protein).  
 GN CRIP9.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 OX NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 282-301.  
 RC STRAIN=cv. W64A;  
 RX MEDLINE=92338851; PubMed=1633495;  
 RA Bass H.W., Webster C., Obrian G.R., Roberts J.K.M., Boston R.S.;  
 RT "A maize ribosome-inactivating protein is controlled by the  
 RT transcriptional activator Opaque-2.";  
 RL Plant Cell 4:225-234(1992).  
 CC -1- FUNCTION: Possesses features of some constitutive defense agent.  
 CC The coordinate Opaque-2-controlled synthesis of this protein and  
 CC the major seed storage proteins (zeins) may provide the  
 CC germinating seedling with both nutritional benefits and protection  
 CC against pathogen invasion of the surrounding endosperm.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of the N-glycosidic bond at one  
 CC specific adenosine on the 28S rRNA.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: Accumulates to high levels in seeds.  
 CC -1- SIMILARITY: Belongs to the ribosome-inactivating protein family.  
 CC Type 1 RIP subfamily.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M83927; AAA33454.1; -.
DR MaizeDB; 30000; -.
DR InterPro; IPR001574; RIP.
DR Pfam; PF00161; RIP; 1.
DR PRINTS; PR00396; SHIGARICIN.
DR PROSITE; PS00275; SHIGA_RICIN; 1.
KW Plant defense; Protein synthesis inhibitor; Hydrolase; Toxin;
KW Multigene family.
FT ACT SITE 208 208 BY SIMILARITY.
FT DOMAIN 183 189 POLY-ALA.
FT DOMAIN 287 295 POLY-ALA.
SQ SEQUENCE 304 AA; 33514 MW; 978789A2DD2DBF3C CRC64;

Alignment Scores:
Pred. No.: 34.1 Length: 304
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x RIP9_MAIZE (1-304)
QY 830 AGCAGCGCGGACGACGACGACG 807
Db 285 ThrThrAlaAlaAlaAlaAla 292

RESULT 16
SOX3_XENLA
ID _SOX3_XENLA STANDARD; PRT; 309 AA.
AC P55863; P40648;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor SOX-3 (SOX-11).
GN SOX3
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]_
RP SEQUENCE FROM N.A.
RA Penzel R.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 51-104 FROM N.A.
RX MEDLINE=92310993; PubMed=1614875;
RA Denny P., Swift S., Brand N., Dabnede N., Barton P., Ashworth A.;
RT "A conserved family of genes related to the testis determining gene,
RT SRY.";
RL Nucleic Acids Res. 20:2887-2887(1992).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 HMG box domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y07542; CAA68828.1; -.

```

```

DR EMBL; X65654; CAA46605.1; -.
DR PIR; S22946; S22946.
DR HSP; Q05066; 1HRV.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_Box; 1.
DR SMART; SM00398; HMG; 1.
DR PROSITE; PS01118; HMG_BOX_2; 1.
KW DNA-binding; Nuclear protein.
FT DNA BIND 40 108 HMG BOX.
FT CONFLICT 85 85 D -> E (IN REF. 2).
SQ SEQUENCE 309 AA; 34034 MW; 89E401E6BB9EAE03 CRC64;

Alignment Scores:
Pred. No.: 34 Length: 309
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.06% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x SOX3_XENLA (1-309)
QY 321 AGCTCCAGACAGTCGCTGCAT 344
Db 279 SerLeuGinAsnSerArgLeuHis 286

RESULT 17
MGB2_HUMAN
ID MGB2_HUMAN STANDARD; PRT; 319 AA.
AC O15479; O75860;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Melanoma-associated antigen B2 (MAGE-B2 antigen) (DSS-AHC critical
DE interval MAGE superfamily 6) (DAM6) (MAGE XP-2).
GN MAGEB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98110575; PubMed=9441743;
RA Lurquin C., de Smet C., Brasseur F., Muscatelli F., Martelange V.,
RA de Plaen E., Brasseur R., Monaco A.P., Boon T.;
RT "Two members of the human MAGEB gene family located in Xp21.3 are
RT expressed in tumors of various histological origins.";
RL Genomics 46:397-408(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=96081328; PubMed=8535061;
RA Dabovic B., Zanaria E., Bardon B., Lisa A., Bordignon C., Russo V.,
RA Mattesi C., Traversari C., Camerino G.;
RT "A family of rapidly evolving genes from the sex reversal critical
RT region in Xp21.";
RL Mamm. Genome 6:571-580(1995).
RN [3]
RP SEQUENCE FROM N.A.
RA McCurdy D.K., Tai L.-Q., Nguyen J., Wang Z., Yang H., Udar N.,
RA Naïem F., Concannon P., Gatti R.A.;
RT "MAGE XP-2: a member of the MAGE gene family isolated from an
RT expression library using systemic lupus erythematosus sera.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Muzny D., Aronson A.D., Adams C., Brundage E., Bunac C., Carvelli K.,
RA Chacko J., Chen J., Di W., Ding Y., Dugan S., Durbin J., Forcum J.,
RA Ganesh R., Garcia C., Goodman M., Gorrell J.H., Haywood M.,
RA Hernandez J., Jackson L., Jin S., Kamal R., Karpathy S., Kovar C.,
RA Leal B., Li Y., Lichtarge O., Liu W., Logan O., Lu J., Ly T.,
RA Martinez C., Oswal G., Perez L., Rashid N.D., Rowland K., Savage L.,
RA Scherer S.E., Shen H., Simon M., Stovall K., Timms K.M., Todd J.,

```

RA Vo Q., Williamson A., Worley K.C., Yu W., Chinault C., Nelson D.,  
RA Gibbs R.A.;  
RA Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahney J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN TESTIS AND PLACENTA, AND IN A  
CC SIGNIFICANT FRACTION OF TUMORS OF VARIOUS HISTOLOGIC TYPES.  
CC -1- SIMILARITY: Contains 1 MAGE domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U91563; AAC23617.1; --  
DR EMBL: AF015766; AAD01565.1; --  
DR EMBL: AC005185; AAD10635.1; --  
DR EMBL: BC026071; AAB26071.1; --  
DR Genew; HGNC:6809; MAGE2.  
DR MIM; 300098; --  
DR InterPro; IPR002190; MAGE.  
DR Pfam; PF01454; MAGE; 1.  
DR PROSITE; PS00838; MAGE; 1.  
KW Antigen; Multigene family.  
FT DOMAIN 111 310 MAGE.  
FT CONFLICT 61 61 X -> E (IN REF. 4).  
SQ SEQUENCE 319 AA; 35276 MW; CF0BD05A232D592D CRC64;

Alignment Scores:  
Pred. No.: 33.8 Length: 319  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x MGB2\_HUMAN (1-319)

QY 830 ACAGCGCGCAGCAGCAGCAGCG 807  
|||||  
Db 67 ThrThraAlaAlaAlaAlaAla 74

RESULT 18  
MAZ\_MESAU STANDARD; PRT; 331 AA.  
ID MAZ\_MESAU  
AC P56670;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Myc-associated zinc finger protein (MAZI) (Purine-binding  
DE transcription factor) (Pur-1) (Fragment).  
GN MAZ.  
OS Mesocricetus auratus (Golden hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Mesocricetus.  
OX NCBI\_TaxID=10036;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=93087555; PubMed=1454839;  
RA Kennedy G.C., Rutter W.J.;  
RT "Pur-1, a zinc-finger protein which binds to purine-rich sequences,  
RT activates an insulin promoter in heterologous cells";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:11498-11502(1992).  
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS, IN VITRO, TO  
CC PURINE-RICH GAGA SITES FOUND IN THE PROMOTER OF MANY GENES  
CC INCLUDING INSULIN I AND II AND ISLET AMYLOID POLYPEPTIDE.  
CC IN VITRO, ACTIVATES TRANSCRIPTION OF RAT INSULIN I PROMOTER  
CC IN BOTH PANCREATIC AND NONPANCREATIC CELLS.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: Ubiquitously expressed.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L06008; -- NOT ANNOTATED\_CDS.  
DR TRANSFAC; T02304; --  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF000096; zf-C2H2; 3.  
DR SMART; SM00355; Znf\_C2H2; 3.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 3.  
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;  
KW Metal-binding; Nuclear protein; Repeat.  
FT NON\_TER 1 1  
FT ZN\_FING 177 199 C2H2-TYPE.  
FT ZN\_FING 266 288 C2H2-TYPE.  
FT ZN\_FING 294 316 C2H2-TYPE.  
FT ZN\_FING 324 >331 C2H2-TYPE.  
FT DOMAIN 83 95 POLY-ALA.  
FT DOMAIN 117 125 POLY-PRO.  
FT DOMAIN 138 141 POLY-ALA.  
FT DOMAIN 144 150 POLY-ALA.  
FT DOMAIN 233 236 POLY-GLY.  
FT NON\_TER 331 331  
SQ SEQUENCE 331 AA; 33374 MW; 0482F1AF43264B32 CRC64;

#### Alignment Scores:

Pred. No.: 33.6 Length: 331  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x MAZ\_MESAU (1-331)

QY 808 GCTGCTGCTGCTGCTGCTGCTGCTT 831  
|||||

Db 90 AlaAlaAlaAlaAlaAlaVal 97

RESULT 19  
IE63 PRVKA  
ID IE63 PRVKA STANDARD; PRT; 361 AA.  
AC Q85232;



```
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Transcriptional regulator IE63 homolog (Protein UL54).
GN UL54.
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=33703;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95363968; PubMed=7637001;
RA Baumeister J., Klupp B.G., Mettenleiter T.C.;
RT "Pseudorabies virus and equine herpesvirus 1 share a nonessential
RT gene which is absent in other herpesviruses and located adjacent to a
RT highly conserved gene cluster.";
RL J. Virol. 69:5560-5567(1995).
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 AND PRV
CC UL54, HSV-2 UL54, EBV-1 5, VZV 4, EBV BMLF1, HCMV UL69, AND HVS-1
CC 57.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X87246; CAA60694.1; -.
DR InterPro; IPR008648; Herpes UL69.
DR Pfam; PF05459; Herpes UL69; 1.
KW Transcription regulation.
SQ SEQUENCE 361 AA; 40451 MW; 44887399D2224870 CRC64;

Alignment Scores:
Pred. No.: 33.2 Length: 361
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x IE63_PRVKA (1-361)

QY 832 CAACGCGCGCAGCAGCAGCAG 809
Db 62 GlnArgArgGlnGlnGlnGln 69

RESULT 20
BRH2_RAT
ID BRH2_RAT STANDARD; PRT; 384 AA.
AC 089181.
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE BarH-like 2 homeobox protein (Bar-class homeodomain protein MBH1)
DE (Homeobox protein B-H1).
GN BARH2 OR MBH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL
RP STAGE.
RC STRAIN=Sprague-Dawley; TISSUE=Head;
RX MEDLINE=98367582; PubMed=9698441;
RA Saito T., Sawamoto K., Okano H., Anderson D.J., Mikoshiba K.;
RT "Mammalian BarH homologue is a potential regulator of neural bHLH
RT genes.";
RL Dev. Biol. 199:216-225(1998).
CC -!- FUNCTION: Potential regulator of neural basic helix-loop-helix

genes. It may down-regulate expression of ASCL1 and, within the
thalamus, up-regulate NGN2, thereby regulating distinct patterns
of neuronal differentiation.
-!- SUBCELLULAR LOCATION: Nuclear.
-!- TISSUE SPECIFICITY: Expressed in the ganglion cell layer of the
retina in the eye and in the ventral zone of the dorsal thalamus
of the CNS.
-!- DEVELOPMENTAL STAGE: Transiently expressed during embryonic
development of the nervous system, detected at 11.5 days old
(E11.5) and declining after E15.5.
-!- SIMILARITY: Belongs to the BAR homeobox family.
-!- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB004056; BAA32474.1; -.
DR HSP; F23441; IFTT.
DR TRANSFAC; T03924; -.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW Homeobox; Transcription regulation; DNA-binding; Nuclear protein;
KW Developmental protein.
FT DOMAIN 85 91 HIS-RICH.
FT DOMAIN 95 98 PRO-RICH.
FT DOMAIN 128 131 ALA-RICH.
FT DOMAIN 327 331 ALA-RICH.
FT DNA_BIND 229 288 HOMEBOX.
SQ SEQUENCE 384 AA; 41463 MW; 0C6CB022ECAF92F CRC64;

Alignment Scores:
Pred. No.: 32.9 Length: 384
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x BRH2_RAT (1-384)

QY 830 ACACGCGCGCAGCAGCAGCAG 807
Db 325 ThrThrAlaAlaAlaAlaAla 332

RESULT 21
SET8_HUMAN
ID SET8_HUMAN STANDARD; PRT; 393 AA.
AC Q9NQR1; Q8TD09;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Histone-lysine N-methyltransferase, H4 lysine-20 specific
DE (EC 2.1.1.43) (Histone H4-K20 methyltransferase) (H4-K20-HMTase)
DE (SET domain-containing protein 8) (PR/SET domain-containing protein
DE 07) (PR/SET07) (PR-Set7).
GN SET8 OR SET07 OR PRSET7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Tain F., Huang S.;
```

RT "A novel PR/SET domain-containing gene, SET07, as a candidate tumor suppressor";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2), SEQUENCE OF 83-103; 109-134; 141-151;  
RP 152-172; 221-230; 245-260; 280-297 AND 350-393, CHARACTERIZATION, AND  
RP MUTAGENESIS OF HIS-340 AND 385-ILE-HIS-393.  
RX MEDLINE=22117191; PubMed=12121615;  
RA Fang J., Feng Q., Ketel C.S., Wang H., Cao R., Xia L.,  
RA Erdjument-Bromage H., Tempst P., Simon J.A., Zhang Y.,  
RT "Purification and functional characterization of SET8, a nucleosomal  
RT histone H4-lysine 20-specific methyltransferase";  
RL Curr. Biol. 12:1086-1099 (2002).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1), SEQUENCE OF 108-131; 220-231 AND  
RP 349-393, CHARACTERIZATION, AND MUTAGENESIS OF ARG-336.  
RC TISSUE=Cervical carcinoma;  
RX MEDLINE=22082172; PubMed=12086618;  
RA Nishio K., Rice J.C., Sarma K., Erdjument-Bromage H., Werner J.,  
RA Wang Y., Chuikov S., Valenzuela P., Tempst P., Steward R., Lis J.T.,  
RA Allis C.D., Reinberg D.;  
RT "PR-Set7 is a nucleosome-specific methyltransferase that modifies  
RT lysine 20 of histone H4 and is associated with silent chromatin";  
RL Mol. Cell 9:1201-1213 (2002).  
CC -1- FUNCTION: Histone methyltransferase. Methylates Lys-20 of histone  
CC H4. H4 Lys-20 methylation represents a specific tag for epigenetic  
CC transcriptional repression. Nucleosomes are preferred as substrate  
CC compared to free histones. May play a role in maintaining silent  
CC chromatin by preventing neighboring acetylation of H4 tail.  
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =  
CC S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.  
CC -1- SUBCELLULAR LOCATION: Nuclear; associates with silent chromatin on  
CC euchromatic arms. No association with constitutive heterochromatin  
CC (By similarity).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event-Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC Name=2;  
CC IsoId=Q9NQR1-1; Sequence=Displayed;  
CC IsoId=Q9NQR1-2; Sequence=VSP\_002226; VSP\_002227;  
CC Note=No experimental confirmation available;  
CC -1- DOMAIN: Although the SET domain contains the active site of  
CC enzymatic activity, both sequences upstream and downstream of the  
CC SET domain are required for methyltransferase activity.  
CC -1- SIMILARITY: BELONGS TO THE HISTONE-LYSINE METHYLTRANSFERASE  
CC FAMILY. PR/SET SUBFAMILY.  
CC -1- SIMILARITY: Contains 1 SET domain.  
CC -1- CAUTION: It is uncertain whether Met-1 or Met-72 is the initiator.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF287261; AAF97812.2; .  
DR EMBL; AY064546; AAL40879.1; ALT\_INIT.  
DR EMBL; AY102937; AAM47033.1; .  
DR MIM; 607240; .  
DR InterPro; IPR001214; SET.  
DR Pfam; PF00856; SET; 1.  
DR SMART; SM00317; SET; 1.  
DR PROSITE; PS50280; SET; 1.  
KW Transferase; Methyltransferase; Chromatin regulator; Nuclear protein;  
KW Coiled coil; Alternative splicing.  
FT DOMAIN 134 163 COILED COIL (POTENTIAL).  
FT DOMAIN 256 382 SET.  
FT DOMAIN 6 67 ALA-RICH.  
FT DOMAIN 29 32 POLY-ARG.  
FT VARSPPLIC 1 41 Missing (in isoform 2).  
FT /FTId=VSP\_002226.

FT VARSPPLIC 42 57 PGRAAGGKMSKPCAVE -> MARGKMSKPRAVEAA (in  
FT isoform 2).  
FT /FTId=VSP\_002227.  
FT R->G: ABOLISHES METHYLTRANSFERASE  
FT ACTIVITY.  
FT H->A: STRONGLY DECREASES  
FT METHYLTRANSFERASE ACTIVITY.  
FT MISSING: ABOLISHES METHYLTRANSFERASE  
FT ACTIVITY.  
FT KG -> RR (IN REF. 1).  
FT D -> A (IN REF. 1).  
FT C -> R (IN REF. 1).  
FT L -> P (IN REF. 1).  
SQ SEQUENCE 393 AA; 42890 MW; 2DCD9B697834B5BD CRC64;  
  
Alignment Scores:  
Pred. No.: 32.8 Length: 393  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 1 Gaps: 0  
  
US-09-729-264-1 (1-1175) x SET8\_HUMAN (1-393)  
  
Qy 808 GCTGCTGCTCTGCGCCGCGTGT 831  
Db 18 AlaAlaAlaAlaAlaAlaValVal 25  
|||||  
|  
  
RESULT 22  
ID FXQ1\_HUMAN STANDARD; PRT; 403 AA.  
AC Q9C009; Q9NS06;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Forkhead box protein Q1 (Hepatocyte nuclear factor 3 forkhead homolog  
DE 1) (HNF-3/forkhead-like protein 1) (HFH-1).  
GN FOXQ1 OR HFH1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.  
RX MEDLINE=216144676; PubMed=11747606;  
RA Bieller A., Pasche B., Frank S., Glaeser B., Kunz J., Watt K.,  
RA Zoll B.;  
RA "Isolation and characterization of the human forkhead gene FOXQ1";  
RL DNA Cell Biol. 20:555-561 (2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21207067; PubMed=11309849;  
RA Hong H.-K., Noveroske J.K., Headon D.J., Liu T., Sy M.S.,  
RA Justice M.J., Chakravarti A.;  
RA "The winged helix/forkhead transcription factor Foxq1 regulates  
RA differentiation of hair in satin mice";  
RL Genesis 29:163-171 (2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Hopkins R.F., Jordan H., Farmer A.A., Rubin G.M., Hong L.,  
RA Dratchenko L., Marusina K., Bontade M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahay J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- TISSUE SPECIFICITY: Expressed predominantly in the stomach,  
 CC trachea, bladder and salivary gland.  
 CC -!- SIMILARITY: Contains 1 fork-head domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF225950; AAK00639.1; -;  
 DR EMBL; AF153341; AAF75586.1; -;  
 DR EMBL; BC053850; AAH53850.1; -;  
 DR HSSP; O63245; 2HFH.  
 DR Genew; HGNC:20951; FOXO1.  
 DR InterPro; IPR001766; TF\_Fork\_head.  
 DR Pfam; PF00250; Fork\_head; 1.  
 DR PRINTS; PR00053; FORKHEAD.  
 DR ProDom; PD000425; TF\_Fork\_head; 1.  
 DR SMART; SM00339; FH; 1.  
 DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE; PS50039; FORK\_HEAD\_3; 1.  
 KW DNA-binding, Nuclear protein; Transcription regulation.  
 FT DNA-BINDING, Nuclear protein; Transcription regulation.  
 FT DOMAIN 119 214 FORK-HEAD.  
 FT DOMAIN 13 103 ALA/GLY-RICH.  
 FT DOMAIN 221 397 PRO-RICH.  
 FT CONFLICT 49 52 NSPA -> KPS (IN REF. 2).  
 FT CONFLICT 60 61 PP -> TQ (IN REF. 2).  
 FT CONFLICT 386 386 S -> L (IN REF. 2).  
 FT CONFLICT 395 395 P -> S (IN REF. 2).  
 SQ SEQUENCE 403 AA; 41491 MW; EB52255AEAC6929B CRC64;  
  
 Alignment Scores:  
 Pred. No.: 32.7 Length: 403  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.06% Indels: 0  
 DB: 1 Gaps: 0  
  
 US-09-729-264-1 (1-1175) x FX01\_HUMAN (1-403)  
 QY 808 GCTGCTGCTGCTGCCCGCTGCT 831  
 DB 80 AlaAlaAlaAlaAlaValVal 87  
  
 RESULT 23  
 RL3 DROME  
 ID - RL3 DROME STANDARD; PRT; 415 AA.  
 AC O16797; Q9VGR4; Q9VGR5; Q9VGR6;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 60S ribosomal protein L3.  
 GN RL3 OR CG4863  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;

RN SEQUENCE FROM N.A. (ISOFORM A).  
 RP MEDLINE=98326317; PubMed=9661671;  
 RX Chan H.Y.E., Zhang Y., Hoheisel J.D., O'Kane C.J.;  
 RA "Identification and characterization of the gene for Drosophila L3  
 RT ribosomal protein";  
 RL Gene 212:119-125 (1998).  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster K.C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalaali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 "The genome sequence of Drosophila melanogaster";  
 Science 287:2185-2195 (2000).  
 [3]  
 RP REVISIONS, AND ALTERNATIVE SPLICING.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.C.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).  
 CC -!- FUNCTION: The L3 protein is a component of the large subunit of  
 CC cytoplasmic ribosomes.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Comment=Experimental confirmation may be lacking for some  
 CC isoforms;  
 CC Name=A; Synonyms=B, E;  
 CC IsoId=O16797-1; Sequence=Displayed;  
 CC Name=D;  
 CC IsoId=O16797-2; Sequence=VSP\_005715, VSP\_005717;

```
CC Name=C;
CC IsoId=016797-3; Sequence=VSP_005714, VSP_005716;
CC -!- SIMILARITY: Belongs to the L3P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF016835; AAC26144.1; -
DR EMBL; AE003690; AAF54609.1; ALT_INIT.
DR EMBL; AE003690; AAF54610.2; -
DR EMBL; AE003690; AAF54611.1; ALT_INIT.
DR EMBL; AE003690; AAF54612.2; -
DR FlyBase; FBgn020910; RPL3.
DR InterPro; IPR000597; Ribosomal L3.
DR InterPro; IPR009000; Translat_factor.
DR Pfam; PF00297; Ribosomal L3; 1.
DR PROSITE; PS00474; RIBOSOMAL L3; 1.
DR Ribosomal protein; Alternative splicing.
DR INIT_MET 0
DR VARSPLIC 121 137 BY SIMILARITY.
DR VARSPLIC 121 137 WYKSKKAPTAKSKWT -> CSSISLURELFKSLNVV
DR (in isoform C).
DR (FTId=VSP_005714).
DR VARSPLIC 123 160 KSKKKAFTKASKWTDDLGGKSIENFRKMLRYCKVIR ->
DR VSEDHVVVLPTFPFVAIFPAPGVHTHKMAADTCFLLESG
DR (in isoform D).
DR (FTId=VSP_005715).
DR Missing (in isoform C).
DR (FTId=VSP_005716).
DR VARSPLIC 161 415 Missing (in isoform D).
DR (FTId=VSP_005717).
DR VARSPLIC 415 AA; 46784 MW; DS5964681FAP8A15 CRC64;
SQ SEQUENCE 415 AA; 46784 MW; DS5964681FAP8A15 CRC64;

Alignment Scores:
Pred. No.: 32.5 Length: 415
Score: 8.00 Matches: 0
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x RL3_DROME (1-415)
QY 830 ACGAGCGCGCAGCAGCAGCAGCG 807
Db 403 ThrThraAlaAlaAlaAlaAla 410

RESULT 24
SOX3 HUMAN
ID SOX3 HUMAN STANDARD; PRT; 443 AA.
AC P41225; P35714;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor SOX-3.
GN SOX3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=94154672; PubMed=8111369;
RA Stevanovic M., Lovell-Badge R., Collignon J., Goodfellow P.N.;
RT "SOX3 is an X-linked gene related to SRY.";
RN Hum. Mol. Genet. 2:2013-2018(1993).
RN [2]
RP SEQUENCE OF 150-203 FROM N.A.
RP MEDLINE=92310993; PubMed=1614875;
RX
```

```
RA Denny P., Swift S., Brand N., Dabhadre N., Barton P., Ashworth A.;
RT "A conserved family of genes related to the testis determining gene,
RT SRY.";
RL Nucleic Acids Res. 20:2887-2887(1992).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 HMG box domain.
CC -!- CAUTION: WAS CALLED SOX-9 BY REF.2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X71135; CAA50465.1; -
DR EMBL; X65665; CAA46616.1; -
DR PIR; I38239; I38239.
DR PIR; S22942; S22942.
DR HSSP; Q05066; 1HRY.
DR TRANSFAC; T04916; -
DR Genew; HGNC:11199; SOX3.
DR MIN; 313430; -
DR GO; GO:0003677; F:DNA binding; TAS.
DR GO; GO:0007417; P:central nervous system development; TAS.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 1.
DR SMART; SM00398; HMG; 1.
DR PROSITE; PS00118; HMG_BOX_2; 1.
DR DNA-binding; Nuclear protein; Transcription regulation.
DR DOMAIN 129 133 POLY-GLY.
DR FT DNA BIND 139 207 HMG BOX.
DR FT DOMAIN 234 248 POLY-ALA.
DR FT DOMAIN 290 294 POLY-PRO.
DR FT DOMAIN 321 327 POLY-ALA.
DR FT DOMAIN 337 344 POLY-ALA.
DR FT DOMAIN 350 361 POLY-ALA.
DR FT CONFLICT 159 159 L -> Q (IN REF. 2).
DR FT CONFLICT 176 176 D -> E (IN REF. 2).
DR FT CONFLICT 202 202 E -> D (IN REF. 2).
DR SEQUENCE 443 AA; 44884 MW; 8031B4EADA52D3B4 CRC64;
SQ SEQUENCE 443 AA; 44884 MW; 8031B4EADA52D3B4 CRC64;

Alignment Scores:
Pred. No.: 32.2 Length: 443
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x SOX3_HUMAN (1-443)
QY 827 ACGCGCGCAGCAGCAGCGGTAT 804
Db 338 ThrAlaAlaAlaAlaAlaAlaTyr 345

RESULT 25
HBB2 WHEAT
ID HBB2 WHEAT STANDARD; PRT; 476 AA.
AC Q41558;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor HBP-1b(c1) (Fragment).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.; AND DNA-BINDING.
RP STRAIN=cv. Horoshirikomugi;
RX
```

```

RX MEDLINE=94193693; PubMed=8144592;
RA Mikami K., Sakamoto A., Iwabuchi M.;
RT "The HBP-1 family of wheat basic/leucine zipper proteins interacts
RL J. Biol. Chem. 269:9974-9985(1994).
CC -!- FUNCTION: Binds to the hexamer motif 5'-ACGTCA-3' of histone gene
CC promoters. Also binds to the hexamer motif in the promoter of the
CC 35S RNA of Cauliflower mosaic virus.
CC -!- SUBUNIT: Binds DNA as a dimer.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the bZIP family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D12921; BAA02305.2; -
CC PIR; C54415; C54415.
CC DR TRANSFAC; T01393; -
CC DR InterPro; IPR004827; TF_bZIP.
CC DR Pfam; PF00170; bZIP; 1.
CC DR SMART; SM00338; BRLZ; 1.
CC DR PROSITE; PS00217; bZIP; 1.
CC DR PROSITE; PS00036; bZIP_BASIC; 1.
CC KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT NON TER 1
FT DNA_BIND 191 210 BASIC MOTIF.
FT DOMAIN 217 231 LEUCINE-ZIPPER.
SQ SEQUENCE 476 AA; 51786 MW; DD4668F6A2932D88 CRC64;

Alignment Scores:
Pred. No.: 31.8 Length: 476
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x HBB2_WHEAT (1-476)
QY 830 AGACGGCGGACGACGACGACG 807
DB 101 ThrThrAlaAlaAlaAlaAla 108

RESULT 26
MAZ MOUSE
ID MAZ_MOUSE STANDARD; PRT; 477 AA.
AC P56671; Q9R1W0;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myc-associated zinc finger protein (MAZI) (Purine-binding
DE transcription factor) (Pur-1).
GN MAZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancratic tumor;
RX MEDLINE=93087555; PubMed=1454839;
RA Kennedy G.C., Rutter W.J.;
RT "Pur-1, a zinc-finger protein which binds to purine-rich sequences,
RT activates an insulin promoter in heterologous cells."
RL Proc. Natl. Acad. Sci. U.S.A. 89:11498-11502(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Liver;

```

```

RX MEDLINE=99192812; PubMed=10092852;
RA Song J., Murakami H., Tsutsui H., Ugai H., Geltlinger C., Murata T.,
RA Matsumura M., Itakura K., Kanazawa I., Sun K., Yokoyama K.K.;
RT "Structural organization and expression of the mouse gene for Pur-1, a
RL highly conserved homolog of the human MAZ gene."
CC Eur. J. Biochem. 259:676-683(1999).
CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR THAT BINDS TO PURINE-RICH
CC GAGA SITES FOUND IN THE PROMOTER OF MANY GENES INCLUDING
CC INSULIN I AND II AND ISLET AMYLOID POLYPEPTIDE.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 6 C2H2-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L04649; -; NOT ANNOTATED CDS.
CC EMBL; AB006360; BAA76280.1; -
CC TRANSFAC; T02303; -
CC DR MGD; MGI:1338623; Maz.
CC DR InterPro; IPR007087; Znf_C2H2.
CC DR Pfam; PF00096; zf-C2H2; 6.
CC DR ProDom; PD000003; Znf_C2H2; 1.
CC DR SMART; SM00355; Znf_C2H2; 6.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
CC DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 5.
CC KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Metal-binding; Nuclear protein; Repeat.
FT ZN_FING 190 212 C2H2-TYPE 1.
FT ZN_FING 279 301 C2H2-TYPE 2.
FT ZN_FING 307 329 C2H2-TYPE 3.
FT ZN_FING 337 360 C2H2-TYPE 4.
FT ZN_FING 366 388 C2H2-TYPE 5.
FT ZN_FING 392 413 C2H2-TYPE 6 (ATYPICAL).
FT DOMAIN 96 108 POLY-ALA.
FT DOMAIN 130 138 POLY-PRO.
FT DOMAIN 151 154 POLY-ALA.
FT DOMAIN 157 163 POLY-ALA.
FT DOMAIN 246 249 POLY-GLY.
FT DOMAIN 435 447 POLY-ALA.
FT CONFLICT 64 64 Q -> QQ (IN REF. 2).
SQ SEQUENCE 477 AA; 48769 MW; F2276C4C7538D2EF CRC64;

Alignment Scores:
Pred. No.: 31.8 Length: 477
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.06% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x MAZ_MOUSE (1-477)
QY 808 GCTGCTGCTGCTGCTGCTGCT 831
DB 103 AlaAlaAlaAlaAlaAlaVal 110

RESULT 27
APM3 YEAST
ID APM3_YEAST STANDARD; PRT; 483 AA.
AC P38153;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Adaptin chain homolog APM3.
GN APM3 OR YKS6 OR YBR288C OR YBR2035.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae.

```

OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE=94378722; PubMed=8091861;  
RA Holmsstroem K., Brandt T., Kallies T.;  
RT "The sequence of a 32,420 bp segment located on the right arm of  
RL chromosome II from *Saccharomyces cerevisiae*.";  
RL Yeast 10:S47-S62(1994).  
CC -!- SIMILARITY: Belongs to the adaptor complexes medium subunit  
CC family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X76053; CAA53651.1; -;  
DR EMBL; Z36157; CAA85253.1; -;  
DR PIR; S44550; S44550.  
DR Germline; 138831; -;  
DR SGD; S0000492; APM3  
DR GO; GO:0006896; P:Golgi to vacuole transport; IMP.  
DR InterPro; IPR008968; AP50.  
DR InterPro; IPR001392; Clathrin med.  
DR Pfam; PF00928; Adap comp sub; 1.  
DR PROSITE; PS00990; CLAT ADAPTOR M.1; 1.  
DR PROSITE; PS00991; CLAT ADAPTOR M.2; 1.  
KW Coated pits.  
SQ SEQUENCE 483 AA; 54879 MW; D96291FF41ACB079 CRC64;  
  
Alignment Scores:  
Pred. No.: 31.8 Length: 483  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 1 Gaps: 0  
  
US-09-729-264-1 (1-1175) x APM3\_YEAST (1-483)  
QY 862 CAACAACGGCAGCAGCGTTCAG 839  
Db 174 GlnGlnArgGlnGlnLeuGln 181  
  
RESULT 28  
MKR4\_HUMAN  
ID MKR4\_HUMAN STANDARD; PRT; 485 AA.  
AC Q13434;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Makorin 4 (Zinc finger protein 127-Xp) (ZNF127-Xp).  
GN MKR4 OR ZNF127L1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hendrich B.D., Longstreet M., Gustashaw K., Nicholls R.D.,  
RA Wallard H.F.;  
RT "An X-linked homologue of the autosomal imprinted gene ZNF127  
RT escapes X chromosome inactivation.";  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
CC -!- SIMILARITY: Contains 4 C3H1-type zinc finger.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; U41315; AAA99070.1; -;  
DR Genew; HGNC:7115; MKRN4.  
DR InterPro; IPR000571; Znf\_CCH.  
DR InterPro; IPR001841; Znf\_fing.  
DR Pfam; PF00097; zf-C3HC4; 1.  
DR Pfam; PF00642; zf-CCH; 4.  
DR SMART; SM00184; RING; 1.  
DR SMART; SM00356; Znf\_C3H1; 4.  
DR PROSITE; PS00518; ZF\_RING\_1; 1.  
DR PROSITE; PS0089; ZF\_RING\_2; 1.  
KW Zinc-finger; Repeat.  
FT ZN\_FING 96 114 C3H1-TYPE 1.  
FT ZN\_FING 125 143 C3H1-TYPE 2.  
FT ZN\_FING 249 267 C3H1-TYPE 3.  
FT DOMAIN 271 298 MAKORIN-TYPE CYS-HIS.  
FT ZN\_FING 316 370 RING-TYPE.  
FT ZN\_FING 405 425 C3H1-TYPE 4.  
SQ SEQUENCE 485 AA; 52909 MW; AE28B962544CEFFE CRC64;  
  
Alignment Scores:  
Pred. No.: 31.7 Length: 485  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 1 Gaps: 0  
  
US-09-729-264-1 (1-1175) x MKR4\_HUMAN (1-485)  
QY 127 ACTGACCGTCTCCAGGGCTGGA 150  
Db 39 ThrAlaProSerProArgAlaGly 46  
  
RESULT 29  
ECR\_BOMMO  
ID ECR\_BOMMO STANDARD; PRT; 606 AA.  
AC P49881;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Ecdysone receptor (Ecdysteroid receptor) (20-hydroxy-ecdysone  
DE receptor) (20E receptor).  
GN ECR OR NR1H1.  
OS Bombyx mori (Silk moth).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Bombycoidea;  
OC Bombycidae; Bombyx.  
OX NCBI\_TaxID=7091;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Myosore; TISSUE=Ovary;  
RX MEDLINE=95360029; PubMed=7633470;  
RA Swevers L., Drevet J.R., Lunke M.D., Iatrou K.;  
RT "The silkmoth homolog of the *Drosophila* ecdysone receptor (B1  
RT isoform): cloning and analysis of expression during follicular cell  
RT differentiation.";  
RL Insect Biochem. Mol. Biol. 25:857-866(1995).  
RN [2]  
RP SEQUENCE OF 61-606 FROM N.A.  
RC STRAIN=Kinsu X Showa; TISSUE=Fat body;  
RA Kaminura M., Tomita S., Fujiwara H.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Receptor for ecdysone. Binds to ecdysone response  
CC elements (ECRES). May play a role in the implementation of the  
CC chorionogenic program at the end of vitellogenesis.  
CC -!- SUBUNIT: Heterodimer of ECR and CFI (Potential).  
CC -!- SUBCELLULAR LOCATION: Nuclear.



DE Polyadenylate-binding protein 1 (Poly(A)-binding protein 1) (PABP 1).  
GN PABP1 OR PABP.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89343997; PubMed=2761544;  
RA Zelus B.D., Giebelhaus D.H., Eib D.W., Kenner K.A., Moon R.T.;  
RT "Expression of the poly(A)-binding protein during development of  
RL xenopus laevis.";  
RL Mol. Cell. Biol. 9:2756-2760(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Gastrula;  
MEDLINE=9106071; PubMed=2209558;  
RA Nietfeld W., Mentzel H., Pieler T.;  
RT "The xenopus laevis poly(A) binding protein is composed of multiple  
RL functionally independent RNA binding domains.";  
RL EMBO J. 9:3699-3705(1990).  
CC -|- FUNCTION: Binds the poly(A) tail of mRNA.  
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -|- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M27072; AAA60936.1; -;  
DR EMBL; X57483; CAA40721.1; -;  
DR PIR; A32323; DNXLPA.  
DR HSP; P11940; ICVJ.  
DR GO; GO:0005737; C:cytoplasm; ISS.  
DR GO; GO:0008143; F:poly(A) binding; ISS.  
DR GO; GO:0016071; P:mRNA metabolism; ISS.  
DR InterPro; IPR002004; PABP/HECT.  
DR InterPro; IPR006515; PABP\_1234.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF00658; PABP; 1.  
DR Pfam; PF00076; rrm; 4.  
DR SMART; SM00517; PolyA; 1.  
DR SMART; SM00360; RRM; 4.  
DR TIGRFAMs; TIGR01628; PABP-1234; 1.  
DR PROSITE; PS50102; RRM; 4.  
DR PROSITE; PS00030; RRM\_RNP\_1; 3.  
DR RNA-binding; Repeat.  
FT DOMAIN 11 89 RNA-BINDING (RRM) 1.  
FT DOMAIN 99 175 RNA-BINDING (RRM) 2.  
FT DOMAIN 191 268 RNA-BINDING (RRM) 3.  
FT DOMAIN 294 370 RNA-BINDING (RRM) 4.  
FT CONFLICT 252 252 Y -> N (IN REF. 2).  
FT CONFLICT 284 284 N -> K (IN REF. 2).  
FT CONFLICT 430 430 R -> S (IN REF. 2).  
FT CONFLICT 603 603 L -> S (IN REF. 2).  
SQ SEQUENCE 633 AA; 70528 MW; 543E0A551DC9E08 CRC64;  
-----  
Alignment Scores:  
Pred. No.: 30.5 Length: 633  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x PAB1\_XENLA (1-633)

QY 830 ACGACGCGCAGCAGCAGCAGCG 807

Db 494 ThrAlaAlaAlaAlaAlaAla 501  
|||||  
RESULT 32  
HMOC DROME  
ID HMOC DROME STANDARD; PRT; 671 AA.  
AC P22810;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Homeotic protein orthodenticle (Ocelliless protein).  
GN ORD OR OC.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91071580; PubMed=1979296;  
RA Finkelstein R., Smouse D., Capaci T.M., Spradling A.C., Perrimon N.;  
RT "The orthodenticle gene encodes a novel homeo domain protein involved  
RL in the development of the Drosophila nervous system and ocular  
RL visual structures.";  
RL Genes Dev. 4:1516-1527(1990).  
CC -|- FUNCTION: Involved in the development of the Drosophila nervous  
CC system and ocular visual structures.  
CC -|- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -|- DEVELOPMENTAL STAGE: Expressed in the anterior region of the  
CC embryo before cellularization and becomes localized to the  
CC procephalic head region following gastrulation.  
CC -|- DOMAIN: Contains multiple repeats consisting of single amino acids  
CC (e.g., Gly, Ser, His, and Asn) and pairs of amino acids (e.g.,  
CC Gly-Val).  
CC -|- SIMILARITY: Belongs to the paired homeobox family.  
CC -|- SIMILARITY: Contains 1 homeobox domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X58983; CAA41732.1; -;  
DR PIR; A35912; A35912.  
DR HSP; P06601; IFJL.  
DR TRANSFAC; T02078; -;  
DR FlyBase; FBgn0004102; oc.  
DR InterPro; IPR001356; Homeobox.  
DR InterPro; IPR007104; Paired\_homeo.  
DR Pfam; PF000046; homeobox; 1.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX 1; 1.  
DR PROSITE; PS00071; HOMEBOX 2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
KW Transcription regulation; Repeat.  
FT DNA\_BIND 73 132 HOMEBOX.  
FT DOMAIN 229 266 2 X 19 AA TANDEM REPEATS.  
FT REPEAT 229 247 1.  
FT REPEAT 248 266 2.  
SQ SEQUENCE 671 AA; 69666 MW; 515B69536E0E9B44 CRC64;  
-----  
Alignment Scores:  
Pred. No.: 30.2 Length: 671  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 1 Gaps: 0





RT "The Groucho/transducin-like enhancer of split transcriptional  
RT repressors interact with the genetically defined amino-terminal  
RL silencing domain of histone H3.";  
J. Biol. Chem. 272:26604-26610(1997).  
[5]  
RN OLIGOMERIZATION, AND INTERACTION WITH HES1.  
RX MEDLINE=9089594; PubMed=9874198;  
RA Grbavac D., Lo R., Liu Y., Stifani S.;  
RT "Transducin-like Enhancer of split 2, a mammalian homologue of  
RT Drosophila Groucho, acts as a transcriptional repressor, interacts  
RT with Hairy/Enhancer of split proteins, and is expressed during  
RT neuronal development.";  
RL Eur. J. Biochem. 258:339-349(1998).  
[6]  
RN INTERACTIONS WITH RUNX1, RUNX3 AND LEF1.  
RX MEDLINE=98426198; PubMed=9751710;  
RA Levanon D., Goldstein R.E., Bernstein Y., Tang H., Goldenberg D.,  
RA Stifani S., Paroush Z., Groner Y.;  
RT "Transcriptional repression by AML1 and LEF-1 is mediated by the  
RT TLE/Groucho corepressors.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:11590-11595(1998).  
[7]  
RN INTERACTIONS WITH UTX AND UTX.  
RX MEDLINE=99072804; PubMed=9854018;  
RA Grbavac D., Lo R., Liu Y., Greenfield A., Stifani S.;  
RT "Groucho/transducin-like enhancer of split (TLE) family members  
RT interact with the yeast transcriptional co-repressor Ssn6 and  
RT mammalian Ssn6-related proteins: implications for evolutionary  
RT conservation of transcription repression mechanisms.";  
RL Biochem. J. 337:13-17(1999).  
[8]  
RN FUNCTION, AND INTERACTION WITH RELA.  
RX MEDLINE=20127929; PubMed=10660609;  
RA Tetsuka T., Uraishi H., Imai H., Ono T., Sonta S.-I., Takahashi N.,  
RA Asamitsu K., Okamoto T.;  
RT "Inhibition of nuclear factor-kappaB-mediated transcription by  
RT association with the amino-terminal enhancer of split, a  
RT Groucho-related protein lacking WD40 repeats.";  
RL J. Biol. Chem. 275:4383-4390(2000).  
[9]  
RN INTERACTION WITH HESX1, AND MUTAGENESIS OF VAL-486; TYR-532; LEU-702  
RX AND SER-715.  
RX MEDLINE=21588048; PubMed=11731482;  
RA Dasen J.S., Martinez Barbera J.-P., Herman T.S., O'Connell S.,  
RA Olson L., Ju B., Tollkuhn J., Baek S.H., Rose D.W., Rosenfeld M.G.;  
RT "Temporal regulation of a paired-like homeodomain repressor/TLE  
RT corepressor complex and a related activator is required for pituitary  
RT organogenesis.";  
RL Genes Dev. 15:3193-3207(2001).  
[10]  
RN SUBCELLULAR LOCATION, AND DEGREE OF PHOSPHORYLATION.  
RX MEDLINE=22384356; PubMed=12397081;  
RA Nuthall H.N., Joachim K., Palaparti A., Stifani S.;  
RT "A role for cell cycle-regulated phosphorylation in Groucho-mediated  
RT transcriptional repression.";  
RL J. Biol. Chem. 277:51049-51057(2002).  
[11]  
RN X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 443-770.  
RX MEDLINE=22053700; PubMed=12057191;  
RA Pickles L.M., Roe S.M., Hemingway E.J., Stifani S., Pearl L.H.;  
RT "Crystal structure of the C-terminal WD40 repeat domain of the human  
RT Groucho/TLE1 transcriptional corepressor.";  
RL Structure 10:751-761(2002).  
CC -1- FUNCTION: Transcriptional corepressor that binds to a number of  
CC transcription factors. Inhibits NF-kappa-B-regulated gene  
CC expression. Inhibits the transcriptional activation mediated by  
CC FOXA2, and by CTNBL1 and TCF family members in Wnt signaling. The  
CC effects of full-length TLE family members may be modulated by  
CC association with dominant-negative AES.  
CC -1- SUBUNIT: Homooligomer and heterooligomer with other family  
CC members. Binds LEF1, RUNX1, RUNX3, FOXA2, UTX, UTY, histone H3,  
CC HESX1, HES1 and the NF-kappa-B subunit RELA.  
CC -1- SUBCELLULAR LOCATION: Nuclear and chromatin-associated, depending

CC on isoforms and phosphorylation status. Hyperphosphorylation  
CC decreases the affinity for nuclear components.  
CC -1- TISSUE SPECIFICITY: In all tissues examined, mostly in brain,  
CC liver and muscle.  
CC -1- PTM: Phosphorylated, probably by CDC2. The degree of  
CC phosphorylation varies throughout the cell cycle, and is highest  
CC at the G2/M transition. Becomes hyperphosphorylated in response to  
CC cell differentiation and interaction with HES1 or RUNX1.  
CC -1- SIMILARITY: Contains 6 WD repeats.  
CC -1- SIMILARITY: Belongs to the WD-repeat Groucho/TLE family.  
CC PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M99435; AAA61192.1; -;  
DR EMBL; BC015747; AAH15747.1; -;  
DR EMBL; BC010100; AAH10100.1; -;  
DR PIR; B56695; B56695.  
DR PDB; 1GXR; 13-JUN-02.  
DR TRANSFAC; T02524; -;  
DR Genew; HGNC:11837; TLE1.  
DR MIM; 600189; -;  
DR GO; GO:0005634; C:nucleus; TAS.  
DR GO; GO:0007275; P:development; TAS.  
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.  
DR GO; GO:0007165; P:signal transduction; TAS.  
DR InterPro; IPR005617; TLE N.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF03920; TLE\_N; 1.  
DR Pfam; PF00400; WD40; 6.  
DR ProDom; PD000018; WD40; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 2.  
DR PROSITE; PS00082; WD\_REPEATS\_2; 2.  
DR PROSITE; PS50294; WD\_REPEATS\_REGION; 2.  
KW Transcription regulation; Repressor; Nuclear protein; Repeat;  
KW WD repeat; Phosphorylation; Wnt signaling pathway; 3D-structure.  
FT DOMAIN 1 131 GLN-RICH.  
FT DOMAIN 132 199 GLY/PRO-RICH.  
FT DOMAIN 200 268 CCN DOMAIN.  
FT DOMAIN 225 228 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 269 449 SER/PRO-RICH.  
FT REPEAT 470 501 WD 1.  
FT REPEAT 528 558 WD 2.  
FT REPEAT 572 602 WD 3.  
FT REPEAT 614 644 WD 4.  
FT REPEAT 696 726 WD 5.  
FT REPEAT 737 767 WD 6.  
FT MOD\_RES 239 239 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
FT MOD\_RES 259 259 PHOSPHORYLATION (BY CDC2) (POTENTIAL).  
FT MOD\_RES 263 263 PHOSPHORYLATION (BY CDC2) (POTENTIAL).  
FT MOD\_RES 267 267 PHOSPHORYLATION (BY CDC2) (POTENTIAL).  
FT MUTAGEN 486 486 V->S: ABOLISHES HESX1 BINDING.  
FT MUTAGEN 532 532 Y->H: ABOLISHES HESX1 BINDING.  
FT MUTAGEN 702 702 L->S: ABOLISHES HESX1 BINDING.  
FT MUTAGEN 715 715 S->P: ABOLISHES HESX1 BINDING.  
FT CONFLICT 407 412 AAAVA -> RGRGR (IN REF. 1).  
FT CONFLICT 464 465 DA -> TP (IN REF. 1).  
SQ SEQUENCE 770 AA; 83200 MW; 695FD1A37410EFE5 CRC64;  
Alignment Scores:  
Pred. No.: 29.6 Length: 770  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 1 Gaps: 0



DR EMBL; U61362; AAB49934.1; --  
 DR EMBL; AY155195; AAN77514.1; --  
 DR EMBL; AY155196; AAN77515.1; --  
 DR EMBL; AY155197; AAN77516.1; --  
 DR EMBL; AY155198; AAN77517.1; --  
 DR EMBL; AY155199; AAN77518.1; --  
 DR EMBL; AY155200; AAN77519.1; --  
 DR EMBL; AK046402; BAC32708.1; --  
 DR EMBL; AK052961; BAC35221.1; --  
 DR EMBL; AK076750; BAC36464.1; --  
 DR EMBL; AK082499; BAC38509.1; --  
 DR MGD; MGI-104636; Tle1  
 DR GO; GO:0005934; C:nucleus; IDA.  
 DR GO; GO:0005515; F:protein binding; IPI.  
 DR GO; GO:0000122; P:negative regulation of transcription from P. .; IDA.  
 DR InterPro; IPR001680; TLE\_N.  
 DR Pfam; PF03920; TLE\_N; 1.  
 DR Pfam; PF00400; WD40; 6.  
 DR ProDom; PD000018; WD40; 1.  
 DR PROSITE; PS00678; WD REPEATS 1; 2.  
 DR PROSITE; PS0082; WD REPEATS 2; 2.  
 DR PROSITE; PS50294; WD REPEATS REGION; 2.  
 DR Transcription regulation; Repressor; Nuclear protein; Repeat;  
 KW WD repeat; Phosphorylation; Wnt signaling pathway;  
 KW Alternative splicing.  
 FT DOMAIN 1 131 GLN-RICH.  
 FT DOMAIN 132 169 GLY/PRO-RICH.  
 FT DOMAIN 200 268 CCN DOMAIN.  
 FT DOMAIN 225 228 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 269 449 SER/PRO-RICH.  
 FT REPEAT 470 501 WD 1.  
 FT REPEAT 528 558 WD 2.  
 FT REPEAT 572 602 WD 3.  
 FT REPEAT 614 644 WD 4.  
 FT REPEAT 696 726 WD 5.  
 FT REPEAT 737 767 WD 6.  
 FT MOD\_RES 237 237 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
 FT MOD\_RES 257 257 PHOSPHORYLATION (BY CDC2) (POTENTIAL).  
 FT MOD\_RES 261 261 PHOSPHORYLATION (BY CDC2) (POTENTIAL).  
 FT MOD\_RES 265 265 PHOSPHORYLATION (BY CDC2) (POTENTIAL).  
 FT VARSPLIC 1 41 MFPSRHPTPHOAGQPFKTPESLDRIKEERQFLQAQYH  
 --> MFTSLCLFPP (in isoform 5).  
 FT VARSPLIC 79 198 /FTid=VSP\_006987.  
 FT VARSPLIC 125 198 Missing (in isoform 3).  
 FT VARSPLIC 127 127 /FTid=VSP\_006988.  
 FT VARSPLIC 127 127 Missing (in isoform 2 and isoform 4).  
 FT VARSPLIC 477 658 /FTid=VSP\_006989.  
 FT VARSPLIC 193 200 Missing (in isoform 8).  
 FT VARSPLIC 201 770 /FTid=VSP\_006991.  
 FT VARSPLIC 649 653 DRESGTSN -> GERPGKPD (in isoform 7 and  
 isoform 8).  
 FT VARSPLIC 654 770 /FTid=VSP\_006992.  
 FT VARSPLIC 48 48 Missing (in isoform 7 and isoform 8).  
 FT VARSPLIC 649 653 RQLQQ -> NKSQY (in isoform 6).  
 FT VARSPLIC 770 770 /FTid=VSP\_006994.  
 FT VARSPLIC 48 48 Missing (in isoform 6).  
 FT VARSPLIC 48 48 /FTid=VSP\_006995.  
 FT CONFLICT 48 48 E -> G (IN REF. 2; AAN77514).  
 Alignment Scores:  
 Pred. No.: 29.6 Length: 770  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.06% Indels: 0  
 DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x TLE1\_MOUSE (1-770)

Qy 808 GCTGCTGCTGCTGCCCGCTGTT 831  
 Db 404 AlaAlaAlaAlaAlaAlaVal 411  
 RESULT 36  
 PKP2 HUMAN  
 ID PKP2 HUMAN STANDARD; PRT; 881 AA.  
 AC Q99959; Q99960;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Plakophilin 2.  
 GN PKP2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RX MEDLINE=9302494; PubMed=10374264;  
 RA Mertens C., Kuhn C., Moll R., Schwetlick I., Franke W.W.;  
 RT "Desmosomal plakophilin 2 as a differentiation marker in normal and  
 malignant tissues.";  
 RL Differentiation 64:277-290(1999).  
 CC -|- FUNCTION: May play a role in junctional plaques.  
 CC -|- SUBCELLULAR LOCATION: Nuclear and associated with desmosomes.  
 CC -|- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=2; Synonyms=5;  
 CC IsoId=Q99959-1; Sequence=Displayed;  
 CC Name=1; Synonyms=A;  
 CC IsoId=Q99959-2; Sequence=VSP\_006736;  
 CC -|- TISSUE SPECIFICITY: WIDELY EXPRESSED. FOUND AT DESMOSOMAL PLAQUES  
 IN SIMPLE AND STRATIFIED EPITHELIA AND IN NON-EPITHELIAL TISSUES  
 SUCH AS MYOCARDIUM AND LYMPH NODE FOLLICLES. IN MOST STRATIFIED  
 EPITHELIA FOUND IN THE DESMOSOMES OF THE BASAL CELL LAYER AND  
 SEEMS TO BE ABSENT FROM SUPRABASAL STRATA.  
 CC -|- SIMILARITY: Belongs to the beta-catenin family.  
 CC -|- SIMILARITY: Contains 8 ARM repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X97675; CAA66264.1; --  
 DR EMBL; X97675; CAA66265.1; --  
 DR Genew; HGNC:9024; PKP2.  
 DR MIM; 602861; --  
 DR GO; GO:0030057; C:desmosome; NAS.  
 DR GO; GO:0016021; C:integral to membrane; TAS.  
 DR GO; GO:0005634; C:nucleus; NAS.  
 DR GO; GO:0005886; C:plasma membrane; TAS.  
 DR GO; GO:0016337; P:cell-cell adhesion; NAS.  
 DR InterPro; IPR008938; ARM.  
 DR InterPro; IPR000225; Armadillo.  
 DR Pfam; PF00514; Armadillo\_seg; 4.  
 DR SMART; SM00185; ARM; 4.  
 DR PROSITE; PS50176; ARM\_REPEAT; 1.  
 KW Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;  
 KW Repeat; Alternative splicing.  
 FT REPEAT 341 383 ARM 1.  
 FT REPEAT 385 424 ARM 2.  
 FT REPEAT

```
FT REPEAT 427 467 ARM 3.
FT REPEAT 571 616 ARM 4.
FT REPEAT 671 711 ARM 5.
FT REPEAT 719 758 ARM 6.
FT REPEAT 807 849 ARM 7.
FT REPEAT 807 849 ARM 8.
FT VARSPLIC 460 503 Missing (in isoform 1).
FT /FTId=VSP 006736.
SQ SEQUENCE 881 AA; 97398 MW; 947838B760EF1D5D CRC64;

Alignment Scores:
Pred. No.: 29 Length: 881
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x PKP2_HUMAN (1-881)
QY 125 GAAGCGAGCCTGGGAGCCCTTCAG 102
Db 700 GluAlaSerLeuGlyAlaLeuGln 707

RESULT 37
A180_MOUSE
ID A180_MOUSE STANDARD; PRT; 901 AA.
AC Q61548; Q61547;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Clathrin coat assembly protein APl80 (Clathrin coat associated protein
DE APl80) (91 kDa synaptosomal-associated protein) (Phosphoprotein Fl-
DE 201).
DN SNAP91.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
EN [1]

SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RX MEDLINE=92300439; PubMed=1607933;
RA Zhou S., Sousa R., Tannery N.H., Lafer E.M.;
RT "Characterization of a novel synapse-specific protein. II. cDNA
RT cloning and sequence analysis of the Fl-20 protein.";
RL J. Neurosci. 12:2144-2155(1992).
CC -!- FUNCTION: Adaptins are components of the adaptor complexes which
CC link clathrin to receptors in coated vesicles. Clathrin-associated
CC protein complexes are believed to interact with the cytoplasmic
CC tails of membrane proteins, leading to their selection and
CC concentration. Binding of APl80 to clathrin triskelia induces
CC their assembly into 60-70 nm coats.
CC -!- SUBCELLULAR LOCATION: Component of the coat surrounding the
CC cytoplasmic face of coated vesicles in the plasma membrane.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q61548-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q61548-2; Sequence=VSP_000172;
CC -!- TISSUE SPECIFICITY: Brain. Associated with the synapses.
CC -!- DEVELOPMENTAL STAGE: Developmentally regulated in a pattern
CC coincident with active synaptogenesis and synaptic maturation.
CC -!- DOMAIN: Possesses a three domain structure: the N-terminal 300
CC residues harbor a clathrin binding site, an acidic middle domain
CC 450 residues, interrupted by an Ala-rich segment, and the C-
CC terminal domain (166 residues).
CC -!- PTM: Phosphorylated.
CC -!- SIMILARITY: Contains 1 epsin N-terminal homology (ENTH) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; M83985; AAA37587.1; -.
CC EMBL; M83985; AAA37586.1; -.
CC PIR; A44825; A44825.
CC MGD; MGI:109132; Snap91.
CC InterPro; IPR001026; ENTH.
CC InterPro; IPR008943; Fl_bind_N.
CC Pfam; PF01417; ENTH; 1.
CC SMART; SMO0273; ENTH; 1.
CC PROSITE; PS00942; ENTH; 1.
CC Coated pits; Alternative splicing; Phosphorylation.
FT DOMAIN 14 145 ENTH.
FT DOMAIN 410 413 POLY-THR.
FT DOMAIN 535 539 POLY-ALA.
FT DOMAIN 547 550 POLY-ALA.
FT DOMAIN 659 664 POLY-SER.
FT DOMAIN 704 710 POLY-SER.
FT VARSPLIC 715 719 Missing (in isoform Short).
FT /FTId=VSP_000172.
SQ SEQUENCE 901 AA; 91851 MW; 24A98FBACE8DB8B1 CRC64;

Alignment Scores:
Pred. No.: 28.9 Length: 901
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x A180_MOUSE (1-901)
QY 833 ACAACGCGCGGCGAGCAGCA 810
Db 532 ThrThrAlaAlaAlaAlaAla 539

RESULT 38
A180_RAT
ID A180_RAT STANDARD; PRT; 915 AA.
AC Q05140;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Clathrin coat assembly protein APl80 (Clathrin coat associated protein
DE APl80) (91 kDa synaptosomal-associated protein).
DN SNAP91.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
EN [1]

SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RX MEDLINE=93178442; PubMed=8440257;
RA Morris S.A., Schroeder S., Plessmann U., Weber K., Ungewickell E.;
RT "Clathrin assembly protein APl80: primary structure, domain
RT organization and identification of a clathrin binding site.";
RL EMBO J. 12:667-675(1993).
CC -!- FUNCTION: Adaptins are components of the adaptor complexes which
CC link clathrin to receptors in coated vesicles. Clathrin-associated
CC protein complexes are believed to interact with the cytoplasmic
CC tails of membrane proteins, leading to their selection and
CC concentration. Binding of APl80 to clathrin triskelia induces
CC their assembly into 60-70 nm coats.
CC -!- SUBCELLULAR LOCATION: Component of the coat surrounding the
CC cytoplasmic face of coated vesicles in the plasma membrane.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q05140-1; Sequence=Displayed;
```

```

CC Name=Short;
CC IsoId=Q05140-2; Sequence=VSP_000173;
CC -! DOMAIN: Possesses a three domain structure: the N-terminal 300
CC residues harbor a clathrin binding site, an acidic middle domain
CC 450 residues, interrupted by an Ala-rich segment, and the C-
CC terminal domain (166 residues).
CC -! PFM: Phosphorylated (166 residues).
CC -! SIMILARITY: Contains 1 epsilon N-terminal homology (ENTH) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68877; CAA48748.1; -.
DR EMBL; X68878; CAA48749.1; -.
DR PIR; S36326; S36326.
DR PIR; S36327; S36327.
DR InterPro; IPR001026; ENTH.
DR InterPro; IPR008943; PI_bind_N.
DR Pfam; PF01417; ENTH; 1.
DR SMART; SM00273; ENTH; 1.
DR PROSITE; PS00942; ENTH; 1.
DR Coated pits; Alternative splicing; Phosphorylation.
FT DOMAIN 14 145 ENTH.
FT DOMAIN 410 413 POLY-THR.
FT DOMAIN 535 539 POLY-ALA.
FT DOMAIN 547 550 POLY-ALA.
FT DOMAIN 678 683 POLY-SER.
FT DOMAIN 723 729 POLY-SER.
FT VARSPLIC 614 632 Missing (in isoform Short).
FT /FTID=VSP_000173.
SQ SEQUENCE 915 AA; 93518 MW; 32EC1B38C5DF8C0 CRC64;

Alignment Scores:
Pred. No.: 28.8 Length: 915
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x A180_RAT (1-915)
QY 833 ACAACGACGGCGCAGCAGCA 810
Db 532 ThrThrThrAlaAlaAlaAlaAla 539

RESULT 39
OVO_DROME OVO_DROME STANDARD; PRT; 1028 AA.
AC P51521; Q9XZU4;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE OVO protein (Shaven baby protein).
GN OVO OR SVB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=95021209; PubMed=7935398;
RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
RT "Multiple products from the shavenbaby-ovo gene region of Drosophila
RT melanogaster: relationship to genetic complexity.";
RL Mol. Cell. Biol. 14:6809-6818(1994).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=91293102; PubMed=1712294;
RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;
RT "The ovo gene of Drosophila encodes a zinc finger protein required
RT for female germ line development.";
RL EMBO J. 10:2259-2266(1991).
CC -! FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
CC LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
CC -! SUBCELLULAR LOCATION: Nuclear (Potential).
CC -! DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIUM AND
CC ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,
CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
CC -! SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U11383; AAB60216.1; -.
DR EMBL; X59772; CAB36921.1; ALT_SEQ.
DR PIR; A56038; A56038.
DR HSPSP; P07248; ZADR.
DR TRANSFAC; T00669; -.
DR FlyBase; FBgn0003028; ovo.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR SMART; SM00355; Znf_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
DR Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 62 66 POLY-ALA.
FT DOMAIN 72 77 POLY-GLY.
FT DOMAIN 80 85 POLY-GLY.
FT DOMAIN 98 108 POLY-GLY.
FT DOMAIN 144 152 POLY-HIS.
FT DOMAIN 153 159 POLY-ASN.
FT DOMAIN 336 339 POLY-GLN.
FT DOMAIN 347 353 POLY-GLN.
FT DOMAIN 357 361 POLY-GLN.
FT DOMAIN 410 414 POLY-GLN.
FT DOMAIN 418 422 POLY-GLN.
FT DOMAIN 426 432 POLY-GLN.
FT DOMAIN 445 453 POLY-GLN.
FT DOMAIN 456 459 POLY-GLN.
FT DOMAIN 466 474 POLY-GLN.
FT DOMAIN 497 517 POLY-ALA.
FT DOMAIN 524 529 POLY-SER.
FT DOMAIN 549 558 POLY-ALA.
FT DOMAIN 639 651 POLY-ALA.
FT DOMAIN 717 725 POLY-ALA.
FT DOMAIN 797 802 POLY-GLN.
FT DOMAIN 820 823 POLY-GLN.
FT DOMAIN 826 832 POLY-GLN.
FT ZN_FING 874 896 C2H2-TYPE 1.
FT ZN_FING 902 924 C2H2-TYPE 2.
FT ZN_FING 930 953 C2H2-TYPE 3.
FT ZN_FING 969 992 C2H2-TYPE 4.
FT CONFLICT 647 647 A -> R (IN REF. 2).
SQ SEQUENCE 1028 AA; 110620 MW; D7068BB2BC0F677 CRC64;

Alignment Scores:
Pred. No.: 28.3 Length: 1028
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

```





RA Waltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelto J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shipakovsky G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).  
RN [2]  
RP SEQUENCE OF 633-836 FROM N.A., AND SUBCELLULAR LOCATION.  
RC STRAIN=968 h90;  
RX MEDLINE=2023889; PubMed=10759889;  
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,  
RA Hiraoka Y.;  
RT "Large-scale screening of intracellular protein localization in living  
RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";  
RL Genes Cells 5:169-190(2000).  
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.  
CC -!- SIMILARITY: Contains 9 TPR repeats.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AL023287; CAAL8877.1; -  
DR EMBL; AB027911; BAB87215.1; -  
DR GeneDB\_SFombe; SPB2386.09; -  
DR InterPro; IPR008941; TPR-like.  
DR InterPro; IPR001440; TPR.  
DR Pfam; PF00515; TPR; 10.  
DR SMART; SM00028; TPR; 9.  
KW Repeat; TPR repeat; Nuclear protein.  
FT DOMAIN 68 204 GLN-RICH.  
FT REPEAT 334 367 TPR 1.  
FT REPEAT 369 401 TPR 2.  
FT REPEAT 402 435 TPR 3.  
FT REPEAT 438 471 TPR 4.  
FT REPEAT 475 508 TPR 5.  
FT REPEAT 512 545 TPR 6.  
FT REPEAT 584 617 TPR 7.  
FT REPEAT 618 651 TPR 8.  
FT REPEAT 653 686 TPR 9.  
FT DOMAIN 893 902 POLY-SER.  
SQ SEQUENCE 1102 AA; 121516 MW; C5258D714C42FE7A CRC64;  
  
Alignment Scores:  
Pred. No.: 28 Length: 1102  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.11% Indels: 0  
DB: 1 Gaps: 0  
  
US-09-729-264-1 (1-1175) x YG49\_SCHPO (1-1102)  
QY 845 GTTGACGACCAACACAGCGCG 822  
D 1038 ValAlaAlaThrThrThrAla 1045  
  
RESULT 42  
MAP4 HUMAN  
ID MAP4\_HUMAN STANDARD; PRT; 1152 AA.  
AC P27816; Q13082; Q9GA76;  
DT 01-AUG-1992 (Rel. 23, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Microtubule-associated protein 4 (MAP 4).  
GN MAP4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=92042100; PubMed=1718985;  
RA West R.R., Tenbarge K.M., Olmsted J.B.;  
RT "A model for microtubule-associated protein 4 structure. Domains  
RT defined by comparisons of human, mouse, and bovine sequences.";  
RL J. Biol. Chem. 266:21886-21896(1991).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=95161404; PubMed=7857940;  
RA Chapin S.J., Lue C.M., Yu M.T., Bulinski J.C.;  
RT "Differential expression of alternatively spliced forms of MAP4: a  
RT repertoire of structurally different microtubule-binding domains.";  
RL Biochemistry 34:2289-2301(1995).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buettow R.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Sapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Bosak S.A., McWeeney P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [4]  
RP SEQUENCE OF 102-1152 FROM N.A. (ISOFORM 1).  
RX MEDLINE=91277031; PubMed=1905296;  
RA Chapin S.J., Bulinski J.C.;  
RT "Non-neuronal 210 x 10(3) Mr microtubule-associated protein (MAP4)  
RT contains a domain homologous to the microtubule-binding domains of  
RT neuronal MAP2 and tau.";  
RL J. Cell Sci. 98:27-36(1991).  
CC -!- FUNCTION: NON-NEURONAL MICROTUBULE-ASSOCIATED PROTEIN. PROMOTES  
CC MICROTUBULE ASSEMBLY.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Comment=Additional isoforms seem to exist;  
CC Name=1;  
CC IsoId=P27816-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=P27816-2; Sequence=VSP\_003200;  
CC -!- PTM: Phosphorylation of the Pro-rich region in the C-terminus  
CC negatively regulates MAP4 activity to promote microtubule  
CC assembly.  
CC -!- SIMILARITY: Contains 4 Tau/MAP repeats.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its



use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M64571; AAA59553.1; -;  
 EMBL; U19727; AAA67361.1; -;  
 EMBL; BC008715; AAH08715.1; -;  
 EMBL; BC012794; AAH12794.1; -;  
 EMBL; BC015149; AAH15149.1; -;  
 PIR; A41206; A33183;  
 Genew; HGNC:6862; MAP4.  
 MIM; 157132; -;  
 GO; GO:0005875; C:Microtubule associated complex; TAS.  
 GO; GO:0005198; F:Structural molecule activity; TAS.  
 InterPro; IPR001084; Tubulin Tau.  
 Pfam; PF00418; tubulin-binding; 4.  
 PROSITE; PS00229; TAU MAP; 4.  
 Microtubule; Repeat; Phosphorylation; Alternative splicing.  
 DOMAIN 248 545 17 X 14 AA TANDEM REPEATS.  
 REPEAT 248 261 1.  
 REPEAT 262 275 2.  
 REPEAT 276 289 3.  
 REPEAT 290 303 4.  
 REPEAT 304 317 5.  
 REPEAT 318 331 6.  
 REPEAT 332 345 7.  
 REPEAT 346 351 8 (INCOMPLETE).  
 REPEAT 352 377 26 RESIDUES 1.  
 REPEAT 378 403 26 RESIDUES 2.  
 REPEAT 408 421 9.  
 REPEAT 422 433 10.  
 REPEAT 434 447 11.  
 REPEAT 448 461 12.  
 REPEAT 462 475 13.  
 REPEAT 476 489 14.  
 REPEAT 490 503 15.  
 REPEAT 504 517 16.  
 REPEAT 532 545 17.  
 TAU/MAP MOTIF 1.  
 TAU/MAP MOTIF 2.  
 TAU/MAP MOTIF 3.  
 TAU/MAP MOTIF 4.  
 Missing (in isoform 2).  
 /FTId=VSP 003200.  
 FT CONFLICT 160 160 A -> R (IN REF. 1).  
 FT CONFLICT 427 427 S -> Y (IN REF. 1).  
 SQ SEQUENCE 1152 AA; 121018 MW; 061A69ACIC067A3A CRC64;

Alignment Scores:  
 Pred. No.: 27.8 Length: 1152  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.11% Indels: 0  
 DB: 1 Gaps: 0  
 FT VARSPLIC 558 730  
 FT CONFLICT 160 160 A -> R (IN REF. 1).  
 FT CONFLICT 427 427 S -> Y (IN REF. 1).  
 SQ SEQUENCE 1152 AA; 121018 MW; 061A69ACIC067A3A CRC64;

US-09-729-264-1 (1-1175) x MAP4\_HUMAN (1-1152)  
 QY 863 ACACACGCGACGACGCGTTCGA 840  
 Db 810 ThrThrThrAlaAlaAlaValAla 817  
 RESULT 43  
 ID1 DPG1\_HUMAN STANDARD; PRT; 1239 AA.  
 AC P54098; Q52515;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE DNA polymerase gamma subunit 1 (EC 2.7.7.7) (Mitochondrial DNA  
 DE polymerase catalytic subunit) (POLG-alpha).  
 GN POLG OR POLG1 OR POLG OR POLGA OR MDPI.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97038687; PubMed=8884268;  
 RA Ropp P.A., Copeland W.C.;  
 RT "Cloning and characterization of the human mitochondrial DNA  
 RT polymerase, DNA polymerase gamma.";  
 RL Genomics 36:449-458(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97186710; PubMed=9034326;  
 RA Lecrenier N.L., van der Bruggen P., Foury F.;  
 RT "Mitochondrial DNA polymerases from yeast to man: a new family of  
 RT polymerases.";  
 RL Gene 185:147-152(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Brain;  
 RA Watanabe T.K., Shimizu F., Nishino N., Fujiwara T., Kanemoto N.,  
 RA Suzuki M., Nakamura Y., Hirai Y., Maekawa H., Takahashi E.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Lymph and Testis;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udgin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP VARIANTS PEO PRO-3; ARG-304; THR-467 AND CYS-955.  
 RX MEDLINE=21324738; PubMed=11431686;  
 RA Van Goethem G., Dermaut B., Lofgren A., Martin J.-J.,  
 RA Van Broeckhoven C.;  
 RT "Mutation of POLG is associated with progressive external  
 RT ophthalmoplegia characterized by mtDNA deletions.";  
 RL Nat. Genet. 28:211-212(2001).  
 CC -!- FUNCTION: Involved in the replication of mitochondrial DNA.  
 CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
 CC + {DNA} (N).  
 CC -!- COFACTOR: Magnesium.  
 CC -!- SUBUNIT: Heterotrimer composed of a catalytic subunit and an  
 CC homodimer of accessory subunits.  
 CC -!- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -!- DISEASE: Defects in POLG are a cause of autosomal dominant  
 CC progressive external ophthalmoplegia with mitochondrial DNA  
 CC deletions (PEO) [MIM:157640]; also known as mitochondrial DNA  
 CC breakage syndrome. Patients with PEO have mitochondrial myopathy,  
 CC progressive external ophthalmoplegia, and other abnormalities  
 CC associated with multiple different deletions of mitochondrial DNA.  
 CC -!- SIMILARITY: Belongs to the DNA polymerase type-A family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; U60325; AAC50712.1; --  
 CC EMBL; X98093; CAA66719.1; --  
 CC EMBL; D84103; BAA12223.1; --  
 CC EMBL; BC042571; AAH42571.1; --  
 CC EMBL; BC050559; AAH50559.1; --  
 CC PIR; G02750; G02750.  
 CC Genew; HGNC:9179; POLG.  
 CC MIM; 174763; --  
 CC MIM; 157640; --  
 CC GO; GO:0005739; C:mitochondrion; TAS.  
 CC GO; GO:0003891; F:delta DNA polymerase activity; TAS.  
 CC GO; GO:0003895; F:gamma DNA-directed DNA polymerase activity; TAS.  
 CC GO; GO:0006261; P:DNA dependent DNA replication; TAS.  
 CC GO; GO:0006259; P:DNA metabolism; TAS.  
 CC InterPro; IPR001098; DNA\_pol.  
 CC InterPro; IPR002297; DNA\_pol.  
 CC Pfam; PF00476; DNA\_pol\_A; 1.  
 CC PRINTS; PR00867; DNAPOLG.  
 CC SMART; SMO0482; POLAC; 1.  
 CC DR PROSITE; PS00447; DNA\_POLYMERASE A; 1.  
 CC DR TRANSFERASE; DNA-directed DNA polymerase; DNA replication;  
 CC KW DNA-binding; Mitochondrion; Magnesium; Disease mutation; Polymorphism.  
 CC FT DOMAIN 43 60  
 CC FT DOMAIN 535 538 POLY-GLU.  
 CC FT VARIANT 3 3  
 CC FT VARIANT 18 18 R -> P (in PEO).  
 CC FT VARIANT 304 304 P -> S (in dBSNP:3087373).  
 CC FT VARIANT 324 324 L -> R (in PEO).  
 CC FT VARIANT 467 467 P -> S (in dBSNP:2307437).  
 CC FT VARIANT 467 467 A -> T (in PEO).  
 CC FT VARIANT 546 546 R -> C (in dBSNP:2307442).  
 CC FT VARIANT 662 662 E -> K (in dBSNP:2307450).  
 CC FT VARIANT 955 955 Y -> C (in PEO).  
 CC FT VARIANT 1142 1142 R -> W (in dBSNP:2307442).  
 CC FT VARIANT 1143 1143 E -> G (in dBSNP:2307441).  
 CC FT VARIANT 1146 1146 R -> C (in dBSNP:2307440).  
 CC FT VARIANT 1236 1236 Q -> H (in dBSNP:3087374).  
 CC FT CONFLICT 55 55 Q -> QQQ (in REF. 3).  
 CC SQ SEQUENCE 1239 AA; 139562 MW; 2D9ECCD75AD6E01E CRC64;

## Alignment Scores:

Pred. No.: 27 5 Length: 1239  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.11% Indels: 0  
 DB: 1 Gaps: 0

US-09-729-264-1 (1-1175) x DPG1\_HUMAN (1-1239)

Qy 832 CAACGACGGCGGACGACGACGACG 809

Db 39 GlnArgArgGlnGlnGlnGln 46

RESULT 44

CCAE\_DISOM

ID CCAC DISOM STANDARD; PRT; 2223 AA.  
 AC P56639;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Probable voltage-dependent R-type calcium channel alpha-1E subunit  
 DE (DOE-1).  
 OS Discopyle ommata (Electric ray).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
 OC Elasmobranchii; Squala; Hynostomalea; Pristigaster; Batoidea;  
 OC Torpediniformes; Narcinoidae; Narcinidae; Discopyle.  
 OC NCBI\_TaxID=7785;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Electric lobe;  
 RC MEDLINE=3248175; PubMed=7683405;  
 RA Horne W.A., Ellinor P.T., Imman I., Zhou M., Tsien R.W., Schwarz T.L.;  
 RT "Molecular diversity of Ca2+ channel alpha 1 subunits from the marine  
 RL ray Discopyle ommata.";  
 CC -I- FUNCTION: The isoform alpha-1E gives rise to R-type calcium  
 CC currents. R-type calcium channels belong to the "high-voltage  
 CC activated" (HVA) group (by similarity).  
 CC -I- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT  
 CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS  
 CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-  
 CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS  
 CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM  
 CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA  
 CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY (BY  
 CC SIMILARITY).  
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -I- TISSUE SPECIFICITY: Expression is higher in the forebrain than in  
 CC the electric lobe.  
 CC -I- DOMAIN: Each of the four internal repeats contains five  
 CC hydrophobic transmembrane segments (S1, S2, S3, S5, S6) and one  
 CC positively charged transmembrane segment (S4). S4 segments  
 CC probably represent the voltage-sensor and are characterized by a  
 CC series of positively charged amino acids at every third position.  
 CC -I- SIMILARITY: Belongs to the calcium channel alpha-1 subunits  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L12531; -, NOT ANNOTATED\_CDS.

PIR; A47447; A47447.

InterPro; IPR001682; Ca/Na pore.

InterPro; IPR002077; Ca channel alpha.

InterPro; IPR002111; Cat channel\_TPrL.

InterPro; IPR005821; Ion trans.

InterPro; IPR005820; M+channel\_nlg.

InterPro; IPR003915; PKD 2.

InterPro; IPR005449; RVDCCalpa1.

Pfam; PF00520; ion\_trans; 4.

PRINTS; PR00167; CACHANNEL.

PRINTS; PR01433; POLYCYSTIN2.

PRINTS; PR01633; RVDCCALPA1.

Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;

Calcium-binding; Phosphorylation.

REPEAT 89 366 I.

REPEAT 478 720 II.

REPEAT 1092 1377 III.

REPEAT 1414 1666 IV.

DOMAIN 1 102 CYTOPLASMIC (POTENTIAL).

DOMAIN 103 121 S1 OF REPEAT I (POTENTIAL).

DOMAIN 122 139 EXTRACELLULAR (POTENTIAL).

```
FT TRANSMEM 140 159 S2 OF REPEAT I (POTENTIAL).
FT DOMAIN 160 171 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 172 189 S3 OF REPEAT I (POTENTIAL).
FT DOMAIN 190 194 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 195 213 S4 OF REPEAT I (POTENTIAL).
FT DOMAIN 214 232 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 233 252 S5 OF REPEAT I (POTENTIAL).
FT DOMAIN 253 338 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 339 363 S6 OF REPEAT I (POTENTIAL).
FT DOMAIN 364 490 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 491 509 S1 OF REPEAT II (POTENTIAL).
FT DOMAIN 510 524 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 525 544 S2 OF REPEAT II (POTENTIAL).
FT DOMAIN 545 552 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 553 571 S3 OF REPEAT II (POTENTIAL).
FT DOMAIN 572 581 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 582 600 S4 OF REPEAT II (POTENTIAL).
FT DOMAIN 601 619 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 620 639 S5 OF REPEAT II (POTENTIAL).
FT DOMAIN 640 692 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 693 717 S6 OF REPEAT II (POTENTIAL).
FT DOMAIN 718 1105 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1106 1124 S1 OF REPEAT III (POTENTIAL).
FT DOMAIN 1125 1140 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1141 1160 S2 OF REPEAT III (POTENTIAL).
FT DOMAIN 1161 1172 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1173 1191 S3 OF REPEAT III (POTENTIAL).
FT DOMAIN 1192 1205 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1206 1224 S4 OF REPEAT III (POTENTIAL).
FT DOMAIN 1225 1243 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1244 1263 S5 OF REPEAT III (POTENTIAL).
FT DOMAIN 1264 1349 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1350 1374 S6 OF REPEAT III (POTENTIAL).
FT DOMAIN 1375 1429 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1430 1448 S1 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1449 1463 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1464 1483 S2 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1484 1491 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1492 1510 S3 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1511 1519 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1520 1538 S4 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1539 1557 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1558 1577 S5 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1578 1638 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1639 1663 S6 OF REPEAT IV (POTENTIAL).
FT DOMAIN 1664 2223 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 33 40 POLY-ALA.
FT DOMAIN 729 735 POLY-SER.
FT DOMAIN 1064 1068 POLY-GLU.
FT DOMAIN 386 403 BINDING TO THE BETA SUBUNIT (BY SIMILARITY).
FT SITE 321 321 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT SITE 671 671 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT SITE 1323 1323 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT SITE 1611 1611 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT CA_BIND 439 450 BY SIMILARITY.
FT MOD_RES 1674 1674 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT CA_BIND 1692 1703 BY SIMILARITY.
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1300 1300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1519 1519 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 2223 AA; 251827 MW; 59722DC03E1CFC5B CRC64;
```

## Alignment Scores:

```
Pred. No.: 25.2 Length: 2223
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.06% Indels: 0
```

```
DB: 1 Gaps: 0
US-09-729-264-1 (1-1175) x CCAE_DISOM (1-2223)
Qy 808 GCTGCTGCTGCTGCGCGCGTGT 831
Db 35 AlaAlaAlaAlaAlaAlaVal 42
RESULT 45
TEGU_HCMVA STANDARD; PRI: 2241 AA.
ID TEGU_HCMVA STANDARD; PRI: 2241 AA.
AC P16785;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Probable large tegument protein.
GN UL48.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_taxID=10360;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
RA Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169.";
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC -!- FUNCTION: Tegument protein.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC HSV-1 24, EBV BFLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17403; CAA35407.1; -
CC PIR; S09811; S09811.
CC InterPro: IPR006928; Herpes_teg_N.
CC Pfam: PF04843; Herpes_teg_N; 1.
CC Capsid assembly.
SQ SEQUENCE 2241 AA; 253222 MW; F8929E9F9D0458E CRC64;
Alignment Scores:
Pred. No.: 25.1 Length: 2241
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.11% Indels: 0
DB: 1 Gaps: 0
US-09-729-264-1 (1-1175) x TEGU_HCMVA (1-2241)
Qy 842 GCAGCCACACACACGCGCGGCA 819
Db 242 AlaAlaThrThrThrAlaAla 249
Search completed: September 18, 2004, 22:54:38
Job time : 35.1337 secs
```



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 18, 2004, 22:44:02 ; Search time 60.6684 Seconds  
(without alignments)  
12221.640 Million cell updates/sec

Title: US-09-729-264-1

Perfect score: 389

Sequence: 1 ctgtctgccatctgaataa.....gtaatacaactgtagtatag 1175

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2013624

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Listing first 135 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTC\_spool\_p/US09729264/runat\_17092004\_155107\_2356/app\_query.fasta\_1.4117  
-DB=SPTRMBL\_25 -QWMT-fascan -SUFFIX=oli.rspt -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -SPART=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=135 -DOCALIGN=200 -THR SCORE=quality -THR\_MIN=0 -ALIGN=45 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09729264@cgn 1 324 @runat\_17092004\_155107\_2356 -NCPU=6 -ICPU=3  
-NO MWAP -LARGEQUERY -NEG-SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

SPTRMBL\_25:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mbc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_rvirus:  
16: sp\_bacteriap:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	Query	Score	Match	Length	DB ID	Description
1	300	77.1	315	4	Q9NS15	Q9ns15 homo sapien

C	2	11	2.9	414	3	Q875P9	Q875p9 saccharomyce
C	3	11	2.9	767	5	Q86KE5	Q86ke5 dictyosteli
C	4	11	2.8	1818	5	Q9VZM1	Q9vzm1 drosophila
C	5	11	2.9	3112	5	Q9NKP1	Q9nkp1 leishmania
C	6	10	2.6	87	10	Q7X7F4	Q7x7f4 oryza sativ
C	7	10	2.6	93	10	Q84Z68	Q84z68 oryza sativ
C	8	10	2.6	102	10	Q94FU5	Q94fu5 hoardeum vul
C	9	10	2.6	170	3	Q871H7	Q871h7 neurospora
C	10	10	2.6	306	10	Q8LH59	Q8lh59 oryza sativ
C	11	10	2.6	390	10	Q406Z5	Q406z5 oryza sativ
C	12	10	2.6	451	10	Q7XQY6	Q7xqy6 oryza sativ
C	13	10	2.6	729	10	Q8H8N4	Q8h8n4 oryza sativ
C	14	10	2.6	2162	5	Q9VQM0	Q9vqm0 drosophila
C	15	10	2.6	2176	5	Q46112	Q46112 drosophila
C	16	10	2.6	3604	5	Q9VVK0	Q9vyk0 drosophila
C	17	9	2.4	41	6	Q28392	Q28392 equus cabal
C	18	9	2.4	79	10	Q8W2Y5	Q8w2y5 oryza sativ
C	19	9	2.4	96	5	Q9VTD0	Q9vtd0 drosophila
C	20	9	2.4	121	11	Q8CEK5	Q8cek5 mus musculu
C	21	9	2.4	127	11	Q8C394	Q8c394 mus musculu
C	22	9	2.3	170	10	Q8H567	Q8h567 oryza sativ
C	23	9	2.4	174	10	Q9LEB6	Q9leb6 petroselinu
C	24	9	2.4	182	5	Q9U9J3	Q9u9j3 toxocara ca
C	25	9	2.4	191	5	Q9W2Z2	Q9w2z2 drosophila
C	26	9	2.3	199	10	Q8H389	Q8h389 oryza sativ
C	27	9	2.4	208	10	Q7Y1F3	Q7y1f3 oryza sativ
C	28	9	2.3	256	10	Q8H558	Q8h558 oryza sativ
C	29	9	2.4	261	12	Q9J8I9	Q9j8i9 spodoptera
C	30	9	2.4	269	5	Q8U9J2	Q8u9j2 toxocara ca
C	31	9	2.3	280	10	Q8S2D6	Q8s2d6 oryza sativ
C	32	9	2.4	285	11	Q9JME9	Q9jme9 mus musculu
C	33	9	2.3	287	3	Q9P5Y9	Q9p5y9 neurospora
C	34	9	2.4	291	5	Q86KV6	Q86kv6 dictyosteli
C	35	9	2.4	310	5	Q97659	Q97659 anopheles g
C	36	9	2.4	327	5	Q76758	Q76758 anopheles g
C	37	9	2.3	327	16	Q93IV2	Q93iv2 streptomyce
C	38	9	2.3	334	2	Q8KNY5	Q8kny5 streptomyce
C	39	9	2.3	343	11	Q7PT22	Q7pt22 rattus norv
C	40	9	2.3	376	2	Q9ZAF5	Q9zaf5 streptomyce
C	41	9	2.4	380	10	Q9SBE2	Q9sbe2 mesembryant
C	42	9	2.4	393	10	Q9ZTF0	Q9ztf0 oryza sativ
C	43	9	2.3	396	16	Q82A13	Q82a13 streptomyce
C	44	9	2.4	402	10	Q9ZRH8	Q9zrh8 oryza sativ
C	45	9	2.3	407	2	Q8KRJ2	Q8krj2 streptomyce
C	46	9	2.3	410	2	Q84AM9	Q84am9 streptomyce
C	47	9	2.4	422	10	Q7XE79	Q7xe79 oryza sativ
C	48	9	2.4	430	10	Q81333	Q81333 mesembryant
C	49	9	2.4	477	10	Q8W0N5	Q8w0n5 oryza sativ
C	50	9	2.4	494	5	Q9VHM0	Q9vhm0 drosophila
C	51	9	2.3	512	10	Q9FW91	Q9fw91 oryza sativ
C	52	9	2.3	512	10	Q7XCU7	Q7xcu7 oryza sativ
C	53	9	2.3	547	10	Q8SAM6	Q8sam6 oryza sativ
C	54	9	2.3	547	10	Q7XGL6	Q7xgl6 oryza sativ
C	55	9	2.4	567	5	Q86BR5	Q86br5 drosophila
C	56	9	2.4	589	5	Q9BLW0	Q9blw0 leishmania
C	57	9	2.4	607	5	Q960E4	Q960e4 drosophila
C	58	9	2.4	639	10	Q84Z02	Q84z02 oryza sativ
C	59	9	2.4	670	5	Q9VEC7	Q9vec7 drosophila
C	60	9	2.4	670	5	Q9NFM7	Q9nfm7 drosophila
C	61	9	2.3	737	16	Q8PQ22	Q8pq22 xanthomonas
C	62	9	2.3	737	16	Q8PD49	Q8pd49 xanthomonas
C	63	9	2.4	912	5	Q9NHC1	Q9nhc1 drosophila
C	64	9	2.4	912	5	Q9NHB9	Q9nhb9 drosophila
C	65	9	2.4	916	5	Q9NHC2	Q9nhc2 drosophila
C	66	9	2.4	916	5	Q9VEX7	Q9vex7 drosophila
C	67	9	2.4	1287	5	Q9NGS5	Q9ngs5 dictyostell
C	68	9	2.4	1351	5	Q9N5F6	Q9n5f6 caenorhabdi
C	69	9	2.4	1403	5	Q9NHN6	Q9nhn6 drosophila
C	70	9	2.4	1403	5	Q8IMM4	Q8imm4 drosophila
C	71	9	2.4	1601	5	Q9V6I9	Q9v6i9 drosophila
C	72	9	2.4	3626	2	Q9F7T9	Q9f7t9 streptomyce
C	73	9	2.4	3972	16	Q9SOR8	Q9sor8 streptomyce
C	74	9	2.4	5532	16	Q9SOR4	Q9sor4 streptomyce



```
QY 579 AATGGGACTTTGACTTGGTGGCTACCTGGAAGACCTGAAGCCCGCAAGTCTGCAACT 638
D 176 AsnGlyThrLeuThrCysValAlaThrTrpYssSerLeuLysAlaAargLysSerAlaThr 195
QY 639 GTAATCTCAGTGTGATTCGGTGTGCCCCAAGACACTGAGGTGGTATTATATCCAGGT 698
D 196 ValAsnLeuThrValLeuArgCysProGlnAspThrGlyGlyGlyLeuAsnLeuProGly 215
QY 659 GTATTATCAAGTTTACCGAGTTTAGTTTTCATGCTACTTGGGCAAGTTGGACTT 758
D 216 ValLeuSerSerLeuProSerLeuGlyPheSerLeuProThrTrpGlyLysValGlyLeu 235
QY 759 GGACTAGCAGGACCATGCTCTGACGCCGACGTGCTACTTACAAATACGCTGCTGCTGC 818
D 236 GlyLeuAlaGlyThrMetLeuLeuThrProThrCysThrLeuThrIleArgCysCys 255
QY 819 TGCCTGCTGCTGTTGTGGTGGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 878
D 256 CysArgArgArgCysCysGlyCysAsnCysCysCysCysCysCysCysCysCysCys 275
QY 879 AAAAGAGGATTCGTTATTCATTTCAAAAGAAATCTGAAAAGAGAGAGACACAAAGAA 938
D 276 LysArgGlyPheArgGlyLeuPheGlnPheGlnLysSerGlnLysSerGlnLysThrAsn 955
QY 939 ACTGAGACAGAAAGTGGAAATGAAATCCGCTACAATTCAGATGAACAAAGACACACA 998
D 296 ThrGlnThrGluSerGlyAsnGluAsnSerGlyTyraSerAspGluGlnLysThrThr 315

RESULT 2
ID Q875P9 PRELIMINARY; PRT; 414 AA.
AC
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE HST1.
OS Saccharomyces kluyveri (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4934;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CB3082;
RX MEDLINE=2242865; PubMed=12594514;
RA Langkjaer R.B., Clifton P.F., Johnston M., Piskur J.;
RT "Yeast genome duplication was followed by asynchronous differentiation
of duplicated genes.";
RL Nature 421:848-852(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CB3082;
RA Langkjaer R.B., Clifton P.F., Johnston M., Piskur J.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY145020; AAC32582.1;
DR GO; GO:0005677; C:chromatin silencing complex; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006342; P:chromatin silencing; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR003000; SIR2.
DR Pfam; PF02146; SIR2; 1.
DR PROSITE; PS03035; SIRTUIN; 1.
SQ SEQUENCE 414 AA; 46122 MW; 6F4D5691E0463F95 CRC64;

Alignment Scores:
Pred. No.: 0.108 Length: 414
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.89% Indels: 0
DB: 3 Gaps: 0

US-09-729-264-1 (1-1175) x Q875P9 (1-414)
```

```
QY 839 GCCACAAACAGCGCGGCGAGCAGCAGCG 807
D 250 AlaThrThrThrThrAlaAlaAlaAlaAla 260

RESULT 3
ID Q86KE5 PRELIMINARY; PRT; 767 AA.
AC
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Turgall B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL Nature 418:79-85(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC116956; AA051109.1;
KW Hypothetical protein.
SQ SEQUENCE 767 AA; 88889 MW; 6FA9653FFC85DD42 CRC64;

Alignment Scores:
Pred. No.: 0.0991 Length: 767
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.89% Indels: 0
DB: 5 Gaps: 0

US-09-729-264-1 (1-1175) x Q86KE5 (1-767)

QY 839 GCCACAAACAGCGCGGCGAGCAGCAGCG 807
D 154 AlaThrThrThrThrAlaAlaAlaAlaAla 164

RESULT 4
ID Q9VZM1 PRELIMINARY; PRT; 1818 AA.
AC
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CG10847 protein.
GN ENC OR CG10847.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
```





```
Q7X7F4
ID Q7X7F4 PRELIMINARY; PRT; 87 AA.
AC Q7X7F4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSUNBA0067G20.23 protein (OSUNB0062B06.2 protein).
GN Oryza sativa (Rice).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,
RA Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,
RA Shao Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,
RA Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,
RA Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z.,
RA Chen L., Fan D.L., Wang Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,
RA Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,
RA Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,
RA Zhang R.Q., Guan J.P., Hong G.F.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Wang Y.J., Lu Y., Li C., Li T.,
RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL606600; CAD40465.1; -
DR EMBL; AL606729; CAE04044.1; -
SQ SEQUENCE 87 AA; 9242 MW; 878D16470DF7C34D CRC64;

Alignment Scores:
Pred. No.: 1.39 Length: 87
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.63% Indels: 0
DB: 10 Gaps: 0

US-09-729-264-1 (1-1175) x Q7X7F4 (1-87)
QY 851 GCAGCAGTTGCAGCACACACACGCGC 822
Db 2 AlaAlaAlaAlaAlaThrThrThrAla 11

RESULT 7
Q84Z68
ID Q84Z68 PRELIMINARY; PRT; 93 AA.
AC Q84Z68;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE P0686C03.15 protein.
GN P0686C03.15.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;

"Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, PAC
clone:P0686C03.15";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004761; BAC56779.1; -
SQ SEQUENCE 93 AA; 9933 MW; C3091D7588087522 CRC64;

Alignment Scores:
Pred. No.: 1.38 Length: 93
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.57% Indels: 0
DB: 10 Gaps: 0

US-09-729-264-1 (1-1175) x Q84Z68 (1-93)
QY 814 GCTGCTGCCGCGTCTGTTGTGCTGCA 843
Db 69 AlaAlaAlaAlaValValValAlaAla 78

RESULT 8
Q94FU5
ID Q94FU5 PRELIMINARY; PRT; 102 AA.
AC Q94FU5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Possible membrane protein LEM1.
GN LEM1.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Morex;
RA Skadsen R.W., Sathish P., Federico M.L., Kaeppler H.F.;
RT Cloning of the promoter for a novel barley gene, Lem1, and its organ-
RT specific promotion of Gfp expression in lemma and palea.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF330255; AAK58425.1; -
SQ SEQUENCE 102 AA; 9706 MW; 96A9BE591C41689F CRC64;

Alignment Scores:
Pred. No.: 1.36 Length: 102
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.63% Indels: 0
DB: 10 Gaps: 0

US-09-729-264-1 (1-1175) x Q94FU5 (1-102)
QY 839 GCCACAAACGACGCGCGCAGCAGCA 810
Db 6 AlaThrThrThrAlaAlaAlaAla 15

RESULT 9
Q871H7
ID Q871H7 PRELIMINARY; PRT; 170 AA.
AC Q871H7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein B16M17.110.
GN B16M17.110.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
```

DE OSJNBb0108J11.5 protein.  
GN OSJNBb0108J11.5.  
OS *Oryza sativa* (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriarthoidae; Oryzae; Oryza.  
 NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,  
 RA Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,  
 RA Shao Y., Sun Y., Gu J.L., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,  
 RA Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,  
 RA Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z.,  
 RA Chen L., Fan D.L., Wang Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,  
 RA Lu T.T., Zhang Y.L., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,  
 RA Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,  
 RA Zhang R.Q., Guan J.P., Hong G.F.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL606618; CAE02913.1; --  
 SQ SEQUENCE 451 AA; 46721 MW; DE97C4C0BDE612F3 CRC64;

Alignment Scores:  
 Pred. No.: 1.11 Length: 451  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.63% Indels: 0  
 DB: 10 Gaps: 0

US-09-729-264-1 (1-1175) x Q7XQY6 (1-451)

QY 836 ACAACACGACGCGCGCAGCAGCAGCA 807  
 Db 10 ThrThrThrThrAlaAlaAlaAlaAla 19

## RESULT 13

Q8H8N4 PRELIMINARY; PRT; 729 AA.

AC O8H8N4;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN OSJNBA0070N04.3  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriarthoidae; Oryzae; Oryza.  
 NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,  
 RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,  
 RA Fadros D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,  
 RA Vanaken S.S., Riedmuller S.B., Uterback T.T., Feldblum T.V.,  
 RA Yang O.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,  
 RA White O., Salzberg S.L., Fraser C.M.;  
 RT "Oryza sativa chromosome 3 BAC OSJNBA0070N04 genomic sequence";  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC091494; AAN65023.1; --  
 DR InterPro; IPR004159; DUF248.  
 DR Pfam; PF03141; DUF248; 1.  
 KW Hypothetical protein.

SQ SEQUENCE 729 AA; 80985 MW; 811D6A22PD606707 CRC64;  
 Alignment Scores:  
 Pred. No.: 1.04 Length: 729  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.63% Indels: 0  
 DB: 10 Gaps: 0

US-09-729-264-1 (1-1175) x Q8H8N4 (1-729)

QY 839 GCCACACACGACGCGCAGCAGCAGCA 810  
 Db 82 AlaThrThrThrAlaAlaAlaAla 91

## RESULT 14

Q9VQM0 PRELIMINARY; PRT; 2162 AA.

AC Q9VQM0;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE CG9660 protein.  
 GN TOC OR BC DNA:LD27161 OR CG9660.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer E.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jallal M.E., Kalush C.D., Kraft G., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Celnik S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jallal M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,

RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome.";  
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RA "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RN SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AEO03581; AAF51145.2; --  
 DR FlyBase; FBgn0015600; toc.  
 DR GO; GO:0004197; P:cysteine-type endopeptidase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000169; SHprot\_acsite.  
 DR PROSITE; PS00639; THIOLE PROTEASE HIS; 1.  
 SQ SEQUENCE 2162 AA; 234123 MW; 4D9A3B7CB45EBDA8 CRC64;

Alignment Scores:  
 Pred. No.: 0.899 Length: 2162  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.63% Indels: 0  
 DB: 5 Gaps: 0

US-09-729-264-1 (1-1175) x Q9VQW0 (1-2162)  
 QY 836 ACAACAACGCGCGGCGAGCAGCAGCG 807  
 Db 967 ThrThrThrAlaAlaAlaAlaAla 976

RESULT 15  
 O46112 PRELIMINARY; PRT; 2176 AA.  
 AC O46112;  
 DT 01-JUN-1998 (TREMELrel. 06, Created)  
 DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE TOUCAN protein.  
 GN TOC OR BCDA:1D27161 OR CG9660.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=9362455; PubMed=9362455;  
 RA Grammont M., Dastugue B., Couderc J.L.;  
 RT "The Drosophila toucan (toc) gene is required in germline cells for  
 RT the somatic cell patterning during oogenesis.";  
 RL Development 124:4917-4926(1997).  
 DR EMBL: Y14157; CAA74574.1; --  
 DR PIR; T13806; T13806  
 DR FlyBase; FBgn0015600; toc.  
 DR GO; GO:0004197; P:cysteine-type endopeptidase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000169; SHprot\_acsite.

DR PROSITE; PS00639; THIOLE PROTEASE HIS; 1.  
 SQ SEQUENCE 2176 AA; 235405 MW; 35ABDE00B49EFC7 CRC64;

Alignment Scores:  
 Pred. No.: 0.898 Length: 2176  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.63% Indels: 0  
 DB: 5 Gaps: 0

US-09-729-264-1 (1-1175) x O46112 (1-2176)  
 QY 836 ACAACAACGCGCGGCGAGCAGCAGCG 807  
 Db 967 ThrThrThrAlaAlaAlaAlaAla 976

RESULT 16  
 Q9VYK0 PRELIMINARY; PRT; 3604 AA.  
 AC Q9VYK0;  
 DT 01-MAY-2000 (TREMELrel. 13, Created)  
 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE CG4013 protein.  
 GN SMR OR CG4013.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RC MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazie R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Stapleton M., Strong R., Sun E.,  
 RA Spier E., Spradling A.C., Stapleton M., Stokowski M.P., Smith T.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).



DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE CG14148 protein.  
GN CG14148.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RX MEDLINE=2019606; PubMed=10731132;  
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Sier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Swirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J., Yao Q.A.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Zhao Q.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
DR EMBL; AB003546; AAF50122.1; -;  
DR FlyBase; FBgn0040821; CG14148.  
SQ SEQUENCE 96 AA; 10884 MW; A17E2752CB1DCA7D CRC64;  
Alignment Scores:  
Pred. No.: 14.3 Length: 96  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.37% Indels: 0  
DB: 5 Gaps: 0  
US-09-729-264-1 (1-1175) x Q9VTD0 (1-96)  
QY 833 ACACGACGGCGCAGCAGCAGCG 807  
Db 20 ThrThrThraAlaAlaAlaAlaAla 28  
RESULT 20  
Q8CEK5  
ID Q8CEK5 PRELIMINARY; PRT; 121 AA.  
Alignment Scores:  
Pred. No.: 14.3 Length: 96  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.37% Indels: 0  
DB: 5 Gaps: 0  
US-09-729-264-1 (1-1175) x Q8CEK5 (1-121)  
QY 833 ACACGACGGCGCAGCAGCAGCG 807  
Db 64 ThrThrThraAlaAlaAlaAlaAla 72  
RESULT 21  
Q8C394  
ID Q8C394 PRELIMINARY; PRT; 127 AA.  
AC Q8C394;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN G630009D10R1K.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CS7BL/6J; TISSUE=Head;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
DR EMBL; AK086583; BAC39697.1; -;  
DR MGD; MGI:12442543; G630009D10R1K.  
KW Hypothetical protein.  
SQ SEQUENCE 127 AA; 13343 MW; 5991BE12D6E45C21 CRC64;  
Alignment Scores:  
Pred. No.: 13.8 Length: 127  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.37% Indels: 0  
DB: 11 Gaps: 0

US-09-729-264-1 (1-1175) x Q8C394 (1-127)  
QY 833 ACAACGCGCGGCGAGCAGCAGCG 807  
Db 22 ThrThrAlaAlaAlaAlaAlaAla 30

RESULT 22  
Q8H567  
ID Q8H567 PRELIMINARY; PRT; 170 AA.  
AC Q8H567;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE OJ1656\_E11.34 protein.  
GN OJ1656\_E11.34.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC  
clone:OJ1656\_E11.34";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RR EMBL: AP003843; BAC24867.1; -;  
DR GO: 0008289; F:lipid binding; IEA.  
DR GO: 0006869; P:lipid transport; IEA.  
DR InterPro: IPR003612; AAI.  
DR InterPro: IPR000528; Plant LTP.  
DR Pfam: PF00234; tryp alpha.amyl. 1.  
DR PRINTS: PR00382; LIPIDNSFER.  
DR SMART: SM00439; AAI; 1.  
SQ SEQUENCE 170 AA; 16355 MW; 93C305FC1DCE7E97 CRC64;

Alignment Scores:  
Pred. No.: 13.2 Length: 170  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 10 Gaps: 0

US-09-729-264-1 (1-1175) x Q8H567 (1-170)  
QY 814 GCTGCTGCGCGCGTGTGTGTGCT 840  
Db 10 AlaAlaAlaAlaAlaAlaAlaAla 18

RESULT 23  
Q9LEB6  
ID Q9LEB6 PRELIMINARY; PRT; 174 AA.  
AC Q9LEB6;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Common plant regulatory factor 7.  
GN CRRF7.  
OS Petroselinum crispum (Parsley) (Petroselinum hortense).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC campanulids; Apiales; Apiaceae; Apioidae; apioid superclade;  
OC Apium Clade; Petroselinum.  
OX NCBI\_TaxID=4043;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ruediger A., Frommeyer H., Naeke C., Wellmer F., Kircher S.,  
RX MEDLINE=21414626; PubMed=11523789;  
RA Schaefer E., Harter K.;

Alignment Scores:  
Pred. No.: 13.2 Length: 170  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 10 Gaps: 0

US-09-729-264-1 (1-1175) x Q9LEB6 (1-174)  
QY 836 ACAACAGCGCGCGGCGAGCAGCA 810  
Db 112 ThrThrAlaAlaAlaAlaAla 120

RESULT 24  
Q9U9J3  
ID Q9U9J3 PRELIMINARY; PRT; 182 AA.  
AC Q9U9J3;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Excretory/secretory mucin MUC-2.  
GN MUC-2.  
OS Toxocara canis (Canine roundworm).  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;  
OC Toxocaridae; Toxocara.  
OX NCBI\_TaxID=6265;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99386876; PubMed=10456930;  
RA Tetteh K.K., Loukas A., Tripp C., Maizels R.M.;  
RT "Identification of abundantly expressed novel and conserved genes from  
the infective larval stage of Toxocara canis by an expressed sequence  
tag strategy.";  
RL Infect. Immun. 67:4771-4779 (1999).  
DR EMBL: AF167707; AAD49339.1; -;  
DR InterPro: IPR003582; ShKT.  
DR Pfam: PF01549; ShTK; 2.  
DR SMART: SM00254; ShKT; 1.  
SQ SEQUENCE 182 AA; 18109 MW; 9DDB9A87F1E46DE9 CRC64;

Alignment Scores:  
Pred. No.: 13.1 Length: 182  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.37% Indels: 0  
DB: 5 Gaps: 0

US-09-729-264-1 (1-1175) x Q9U9J3 (1-182)  
QY 842 GCAGCCACACACGCGCGGCGAGCA 816  
Db 28 AlaAlaThrThrAlaAlaAla 36

RESULT 25

QW222  
ID Q9W222 PRELIMINARY; PRT; 191 AA.  
AC Q9W222;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE CG15314 protein.  
GN CG15314.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananikian P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer J.R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA April J.P., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Balow R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beson K.Y., Benos P.V., Berman B.P., Bhargava D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davidson L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskey D.R., Pacieb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svarksas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.M.,  
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195 (2000).  
DR EMBL; AE003449; AAF46545.1; -  
DR FlyBase; FBgn00301172; CG15314.  
SQ SEQUENCE 191 AA; 19870 MW; C596077F70696170B CRC64;

Alignment Scores:  
Pred. No.: 13 Length: 191  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.37% Indels: 0  
DB: 5 Gaps: 0

US-09-729-264-1 (1-1175) x Q9W222 (1-191)

Oy 842 GCAGCCACAACACGCGCGCAGCA 816  
Db 89 AlalaThrThrThrAlaAlaAla 97



Pred. No.: 12.9 Length: 208  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 2.37%  
 DB: 10  
 Gaps: 0

US-09-729-264-1 (1-1175) x Q7YIF3 (1-208)

QY 833 ACAACAGCGCGCAGCAGCAGCAGCG 807

Db 50 ThrThrThrAlaAlaAlaAlaAla 58

RESULT 28

Q8H58 PRELIMINARY; PRT; 256 AA.

ID Q8H58

AC Q8H58

DT 01-MAR-2003 (TREMBLrel. 23, Created)

DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE QJ1714\_H10.16 protein.

GN QJ1714\_H10.16

OS Oryza sativa (japonica cultivar-group)

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzaceae; Oryza.

OX NCBI\_TaxID=39947;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Nipponbare;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC

clone:QJ1714\_H10.16";

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AP003847; BAC15940.1; -

DR GO; GO:0005489; F:electron transporter activity; IEA.

DR GO; GO:0006118; P:electron transport; IEA.

DR InterPro; IPR002109; Glutaredoxin.

DR Pfam; PF00462; Glutaredoxin; 1.

SQ SEQUENCE 256 AA; 27068 MW; F6225BCC86B17F57 CRC64;

Alignment Scores:

Pred. No.: 12.5 Length: 256  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 2.31%  
 DB: 10  
 Gaps: 0

US-09-729-264-1 (1-1175) x Q8H58 (1-256)

QY 814 GCTGCTGCCGCGTGGTGGTGGCT 840

Db 34 AlaAlaAlaAlaValValValAla 42

RESULT 29

ID Q9J819

AC Q9J819

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

DE ORF118.

OS Spodoptera exigua nucleopolyhedrovirus.

OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;

OC Nucleopolyhedrovirus.

OX NCBI\_TaxID=10454;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20036646; PubMed=10567663;

RA Ickel W.F., van Strien E.A., Heldens J.G., Broer R., Zuidema D.,

RA Goldbach R.W., Vlask J.M.;

RT "Sequence and organization of the spodoptera exigua multicapsid

RT nucleopolyhedrovirus genome.";  
 RL J. Gen. Virol. 80:3289-3304(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ickel W.F.J., van Strien E.A., Heldens J.G.M., Broer R., Zuidema D.,  
 RA Goldbach R.W., Vlask J.M.;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF169823; AAF33647.1; -

SQ SEQUENCE 261 AA; 31359 MW; A63B13A1F7FA7457 CRC64;

Alignment Scores:

Pred. No.: 12.5 Length: 261  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 2.37%  
 DB: 12  
 Gaps: 0

US-09-729-264-1 (1-1175) x Q9J819 (1-261)

QY 835 CAACACGACGCGCAGCAGCAGCAG 809

Db 229 GlnGlnArgArgGlnGlnGln 237

RESULT 30

Q9U9J2

ID Q9U9J2

AC Q9U9J2

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Excretory/secretory mucin MUC-3.

GN MUC-3.

OS Toxocara canis (Canine roundworm).

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;

OC Toxocaridae; Toxocara.

OX NCBI\_TaxID=6265;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99386876; PubMed=10456930;

RA Tetteh K.K., Loukas A., Tripp C., Maizels R.M.;

RT "Identification of abundantly expressed novel and conserved genes from

the infective larval stage of Toxocara canis by an expressed sequence

tag strategy";

RL Infect. Immun. 67:4771-4779(1999).

DR EMBL; AF167708; AAD49340.1; -

DR GO; GO:0003824; F:catalytic activity; IEA.

DR InterPro; IPR000385; MoaA\_NiFB\_Pqge.

DR Pfam; PF01549; ShTK; 4.

DR SMART; SM00254; ShTK; 2.

DR PROSITE; PS01305; MOA\_NiFB\_POQE; 1.

SQ SEQUENCE 269 AA; 27940 MW; 4426376C37867E73 CRC64;

Alignment Scores:

Pred. No.: 12.4 Length: 269  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 2.37%  
 DB: 5  
 Gaps: 0

US-09-729-264-1 (1-1175) x Q9U9J2 (1-269)

QY 842 GCAGCCACAAACGCGCGCGCAGCA 816

Db 128 AlaAlaThrThrThrAlaAlaAla 136

RESULT 31

Q8S2D6

ID Q8S2D6

AC Q8S2D6

DT 01-JUN-2002 (TREMBLrel. 21, Created)

DR	SMART: SM00409; IG: 1.
DR	PROSITE: PS0835; IG_LIKE: 1.
KW	Hypothetical protein.
SQ	SEQUENCE 285 AA; 30159 MW; 1CED2C02E2C93BF1 CRC64;
Alignment Scores:	
Pred. No.:	12.3 Length: 285
Score:	9.00 Matches: 9
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	2.37% Indels: 0
DB:	11 Gaps: 0
US-09-729-264-1 (1-1175) x QJ9JME9 (1-285)	
Qy	833 ACACGAGCGGGCCAGCAGCAGCG 807
Dd	228 ThrThrAlaAlaAlaAlaVal 236
RESULT 33	
QPSY9	
ID	PRELIMINARY; PRT; 287 AA.
AC	Q9P5Y9
DT	01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Hypothetical protein.
CN	B208.280.
OS	Neurospora crassa.
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariaceae; Neurospora
OC	Sordariomycetidae; Sordariales; Sordariaceae; Neurospora
OX	NCBI_TaxID=5141;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Schulte U., Aign V., Hoheisel J., Brandt P., Fairmann
RA	Nyakatura G., Wewes H.W., Mannhaupt G.;
RL	Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases
RN	[2]
RP	SEQUENCE FROM N.A.
RA	German Neurospora genome project;
RL	Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases
DR	EMBL; AL355930; CAB91381.2; --
DR	FIR; T49323; T49329.
DR	GO; GO:0016020; C:membrane; IEA.
DR	GO; GO:0005215; F:transporter activity; IEA.
DR	GO; GO:0006810; P:transport; IEA.
DR	InterPro; IPR000425; MIP.
DR	Protein; IPRO00626; Ubiquitin.
DR	PROSITE; PS00221; MIP; 1.
DR	PROSITE; PS0053; UBQUITIN_2; 1.
KW	Hypochemical protein.
SQ	SEQUENCE 287 AA; 29805 MW; C9C447F7975B8BC3 CRC64;
Alignment Scores:	
Pred. No.:	12.3 Length: 287
Score:	9.00 Matches: 9
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	2.31% Indels: 0
DB:	3 Gaps: 0
US-09-729-264-1 (1-1175) x Q9P5Y9 (1-287)	
Qy	808 GCTGTCTCTGTCGCCGCNGTGT 834
Dd	92 AlaAlaAlaAlaAlaValval 100
RESULT 34	
Q86KV6	
ID	PRELIMINARY; PRT; 291 AA.
AC	Q86KV6
DT	01-JUN-2003 (TrEMBLrel. 24, Created)
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4;  
 RX MEDLINE=22092622; PubMed=12097910;  
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,  
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
 RA Tungal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.,  
 RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.",  
 RL Nature 418:79-85(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4;  
 RC STRAIN=AX4;  
 RA Baumgart C.;  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, ACL16920; AAC050883.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 291 AA; 33669 MW; A871B67BA370F9DB CRC64;

Alignment Scores:  
 Pred. No.: 12.3 Length: 291  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.37% Indels: 0  
 DB: 5 Gaps: 0

US-09-729-264-1 (1-1175) x Q86KV6 (1-291)  
 QY 836 ACAACACGCGCGCAGCAGCA 810  
 Db 134 ThrThrAlaAlaAlaAla 142

RESULT 35  
 O76759 PRELIMINARY; PRT; 310 AA.  
 ID O76759;  
 AC O76759;  
 DT 01-NOV-1998 (TReMBLrel. 08, Created)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 DE Ultrabithorax homeotic protein IVA.  
 GN UBX.  
 OS Anopheles gambiae (African malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.  
 OX NCBI\_TaxID=7165;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=G3;  
 RA Devenport M.P., Eggleston P.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL; AF080563; AAC31943.1; -;  
 DR HSSP; P02834; 1B81.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR001827; Antennapedia.  
 DR InterPro; IPR001356; Homeobox.  
 DR Pfam; PF00046; homeobox; 1.  
 DR PRINTS; PR00025; ANTENNAPEDIA.  
 DR PRINTS; PR00024; HOMEBOX.  
 DR ProDom; PD000010; Homeobox; 1.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00032; ANTENNAPEDIA; 1.  
 DR PROSITE; PS00027; HOMEBOX 1; 1.  
 DR PROSITE; PS50071; HOMEBOX 2; 1.  
 KW DNA-binding; Homeobox; Nuclear protein.  
 RN [1]  
 RP SEQUENCE 310 AA; 33203 MW; 65B2A9940E62D085 CRC64;

Alignment Scores:  
 Pred. No.: 12.1 Length: 327  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.37% Indels: 0  
 DB: 5 Gaps: 0

US-09-729-264-1 (1-1175) x O76758 (1-327)  
 QY 830 ACACGCGCGCAGCAGCGGTAT 804  
 Db 32 ThrThrAlaAlaAlaAlaTyr 40

RESULT 37  
 Q93IV2 PRELIMINARY; PRT; 327 AA.  
 ID Q93IV2  
 AC Q93IV2;  
 DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Hypothetical protein SC01318.  
GN SC01318 OR SCBAC36F5.29.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RX MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2)."; (2002).  
RL Nature 417:141-147 (2002).  
DR EMBL; AL939108; CAC42865.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 327 AA; 33244 MW; B595F20C86A32597 CRC64;

Alignment Scores:  
Pred. No.: 12.1 Length: 327  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 16 Gaps: 0

US-09-729-264-1 (1-1175) x Q93IV2 (1-327)

QY 817 GCTGCCCGCTGTTGTGTGGCTGCA 843  
|||||  
Db 74 ALaLaLaValValValAlaLa 82

RESULT 38  
Q8KNY5 PRELIMINARY; PRT; 334 AA.  
ID Q8KNY5  
AC Q8KNY5; 2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DE Transglutaminase (Fragment).  
OS Streptomyces mobaraensis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=35621;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 27441;  
RA Jiang S.-T., Tzeng S.-S., Wu W.-T., Chen G.-H.;  
PL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY129279; AAN01353.1; -.  
FT NON TER 1  
SQ SEQUENCE 334 AA; 38168 MW; 8C2356C48FC2CB35 CRC64;

Alignment Scores:  
Pred. No.: 12.1 Length: 334  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x Q8KNY5 (1-334)

QY 356 CTTACCGTCCAGTTATGGGAGAGCT 382  
|||||

Db 22 ProTyrArgProSerTyrGlyArgala 30

RESULT 39  
Q7TP22 PRELIMINARY; PRT; 343 AA.  
ID Q7TP22  
AC Q7TP22; 2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DE Cbl-725  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Xu C.S., Li W.Q., Li Y.C., Yuan J.Y., Yang K.J., Yan H.M., Chang C.F.,  
RA Zhao L.F., Ma H., Wang L., Wang S.F., Han H.P., Wang G.P., Chai L.Q.,  
RA Shi J.B., Rahman S., Wang Q.N., Zhang J.B.;  
RT "Liver regeneration after PH.";  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY325233; AAP92634.1; -.  
SQ SEQUENCE 343 AA; 37951 MW; E72E38AE1FA7719E CRC64;

Alignment Scores:  
Pred. No.: 12 Length: 343  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0  
DB: 11 Gaps: 0

US-09-729-264-1 (1-1175) x Q7TP22 (1-343)

QY 144 GGCTGGAAGCTCATGCTGGCTCTC 170  
|||||  
Db 60 GlyTyrPhysLeulleMetTrpAlaLeu 68

RESULT 40  
Q9ZAF5 PRELIMINARY; PRT; 376 AA.  
ID Q9ZAF5  
AC Q9ZAF5; 1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Transglutaminase (EC 2.3.2.13) (Fragment).  
OS Streptomyces mobaraensis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=35621;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSMZ;  
RX MEDLINE=99053680; PubMed=9839945;  
RA Pasternack R., Dorsch S., Otterbach J.T., Robenek I.R., Wolf S.,  
RA Fuchsbaue H.L.;  
RT "Bacterial pro-transglutaminase from Streptococcus mobaraensis :  
RT purification, characterisation and sequence of the zymogen.";  
RL Eur. J. Biochem. 257:570-576 (1998).  
DR EMBL; Y18315; CAA77128.1; -.  
DR GO; GO:0008415; F:acetyltransferase activity; IEA.  
DR DR GO; GO:0003810; F:protein-glutamine gamma-glutamyltransferase. . .; IEA.  
DR DR GO; GO:0016740; F:transferase activity; IEA.  
KW Acyltransferase; Transferase.  
FT NON TER 1  
SQ SEQUENCE 376 AA; 42445 MW; 15FE7474DE3771B9 CRC64;

Alignment Scores:  
Pred. No.: 11.9 Length: 376  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.31% Indels: 0

DB: 2 Gaps: 0  
 US-09-729-264-1 (1-1175) x Q9ZAF5 (1-376)  
 QY 356 CCTTACCGTCACAGTTATGGGAGGCT 382  
 Db 64 ProTyrArgProSerTyrGlyArgAla 72  
 RESULT 41  
 Q9SBE2  
 ID Q9SBE2 PRELIMINARY; PRT; 380 AA.  
 AC Q9SBE2;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Stomal L-ascorbate peroxidase (EC 1.11.1.11)  
 OS Mesembryanthemum crystallinum (Common ice plant)  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Caryophyllales; Alzooaceae; Mesembryanthemum.  
 OX NCBI\_TaxID=3544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Michalowski C.B., Quigley-Landreau F., Bohnert H.J.;  
 RT "A stomal ascorbate peroxidase from the common ice plant."  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF069316; AAC19394.1; -;  
 DR HSSP; P48534; IAPX.  
 DR GO; GO:0016688; F:L-ascorbate peroxidase activity; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0004601; F:peroxidase activity; IEA.  
 DR GO; GO:0006979; P:response to oxidative stress; IEA.  
 DR InterPro; IPR002016; Peroxidase.  
 DR Pfam; PF00141; peroxidase; 1.  
 DR PRINTS; PR00458; PEROXIDASE.  
 DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
 DR PROSITE; PS00873; PEROXIDASE\_4; 1.  
 KW Oxidoreductase; Peroxidase.  
 SQ SEQUENCE 380 AA; 40927 MW; 3AB819E6F89DAC5A CRC64;  
 Alignment Scores:  
 Pred. No.: 11.9 Length: 380  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.37% Indels: 0  
 DB: 10 Gaps: 0  
 US-09-729-264-1 (1-1175) x Q9SBE2 (1-380)  
 QY 833 ACAACGCGCGGCGAGCAGCAGCG 807  
 Db 11 ThrThrThrAlaAlaAlaAlaAla 19  
 RESULT 42  
 Q9ZTP0  
 ID Q9ZTP0 PRELIMINARY; PRT; 393 AA.  
 AC Q9ZTP0;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN OSE705.  
 OS Oryza sativa (Rice).  
 CC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 CC Eriatodeae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chen P.W., Chen L.J.;  
 RC STEAIN=Lomello;  
 RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF049348; AAD02494.1; -;  
 DR Gramene; Q9ZTP0; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 393 AA; 45258 MW; DBD01934BA2F9E95 CRC64;  
 Alignment Scores:  
 Pred. No.: 11.8 Length: 393  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.37% Indels: 0  
 DB: 10 Gaps: 0  
 US-09-729-264-1 (1-1175) x Q9ZTP0 (1-393)  
 QY 836 ACAACGCGCGGCGAGCAGCAGCA 810  
 Db 339 ThrThrThrAlaAlaAlaAlaAla 347  
 RESULT 43  
 Q82A13  
 ID Q82A13 PRELIMINARY; PRT; 396 AA.  
 AC Q82A13;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN SAV6246.  
 OS Streptomyces avermitilis.  
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 CC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=33903;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
 RX MEDLINE=21477403; PubMed=11572948;  
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
 RT "Genome sequence of an industrial microorganism Streptomyces  
 avermitilis: deducing the ability of producing secondary  
 metabolites."  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
 RX MEDLINE=22608306; PubMed=12692562;  
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
 RA Sakaki Y., Hattori M., Omura S.;  
 RT "Complete genome sequence and comparative analysis of the industrial  
 microorganism Streptomyces avermitilis."  
 RL Nat. Biotechnol. 21:526-531(2003).  
 DR EMBL; AP005046; BAC73957.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 396 AA; 42979 MW; 689C6599AD28DAEC CRC64;  
 Alignment Scores:  
 Pred. No.: 11.8 Length: 396  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.31% Indels: 0  
 DB: 16 Gaps: 0  
 US-09-729-264-1 (1-1175) x Q82A13 (1-396)  
 QY 829 GTTGTGCTGGCTGCACTGCTGCTGCC 855  
 Db 9 ValValValAlaAlaThrAlaAlaAla 17  
 RESULT 44  
 Q9ZRH8  
 ID Q9ZRH8 PRELIMINARY; PRT; 402 AA.

AC Q9ZRH8;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Early embryogenesis protein.  
 GN OSE362.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Tainung 67;  
 RA Tseng M.J., Wang C.S., Hsu H.R.;  
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U25969; AAD10370.1; -.  
 DR Gramene; Q9ZRH8; -.  
 SQ SEQUENCE 402 AA; 44252 MW; F03C8948F840060 CRC64;

Alignment Scores:  
 Pred. No.: 11.8 Length: 402  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.37% Indels: 0  
 DB: 10 Gaps: 0

US-09-729-264-1 (1-1175) x Q9ZRH8 (1-402)

QY 836 ACAACAGCAGCGCGCAGCAGCA 810

DB 337 ThrThrThrAlaAlaAlaAlaAla 345

RESULT 45

Q8KRJ2 PRELIMINARY; PRT; 407 AA.  
 AC Q8KRJ2;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Transglutaminase precursor.  
 OS Streptomyces mobaraensis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=35621;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IFO13819;  
 RA Kikuchi Y., Date M., Yokoyama K.-I., Umezawa Y., Matsui H.;  
 RT "Secretion of active form transglutaminase of Streptovorticillium  
 RT mobaraense in Corynebacterium glutamicum: Processing of pro-domain  
 RT with co-secreted subtilisin-like protease from Streptomyces  
 RT albogriseolus.";  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF531437; AAM95951.1; -.  
 DR PDB; 1IU4; 04-MAR-03.  
 KW Signal.  
 FT SIGNAL 1 31 POTENTIAL.  
 FT CHAIN 77 407 TRANSGLUTAMINASE.  
 SQ SEQUENCE 407 AA; 45684 MW; 10F7FA04EA52DF4 CRC64;

Alignment Scores:  
 Pred. No.: 11.8 Length: 407  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.31% Indels: 0  
 DB: 2 Gaps: 0

US-09-729-264-1 (1-1175) x Q8KRJ2 (1-407)

QY 356 CCTTACCGTCACGTTATGGAGAGCT 382

Db 95 ProTyrArgProSerTyrGlyArgAla 103  
 Search completed: September 18, 2004, 23:00:58  
 Job time : 81.6684 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 18, 2004, 22:47:02 ; Search time 13.8543 Seconds  
(without alignments)  
8704.746 Million cell updates/sec

Title: US-09-729-264-3

Perfect score: 388

Sequence: 1 agtgatcatggtggcaggag.....gtaatacaactgtagtatag 1168

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Word size: 0

Total number of hits satisfying chosen parameters: 400536

Minimum DB seq length: 25

Maximum DB seq length: 2000000000

Post-processing: Listing first 135 summaries

Command line parameters:

-MODEL=frame+hzp.model -DEV=xlp  
-DB=/cgcn2\_1/USPTO\_spool\_p/US09729264/runat 17092004 155108 2401/app query.fasta\_1.4117  
-QB=Issued Patents AA -QFMT=fastan -SUFFIX=oli.rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=135 -DOCLALIGN=200 -THR\_SCORE=quality -THR\_MIN=0 -ALIGN=45 -MODE=LOCAL  
-OUTFMT=ptio -NORM=ext -HAPSIZE=500 -MINLEN=25 -MAXLEN=2000000000  
-USER=US09729264@CGN 1 1 74 @runat 17092004 155108 2401 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:  
1: /cgcn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgcn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgcn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgcn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgcn2\_6/ptodata/2/iaa/6C\_COMB.pep:\*  
6: /cgcn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	2-3	269	4	US-09-328-352-7889
2	9	2-3	331	1	US-08-136-993-1
3	9	2-3	331	3	US-09-109-063-1
4	9	2-3	331	3	US-08-793-426A-3
5	9	2-3	331	3	US-09-294-565-3
6	9	2-3	331	4	US-09-448-310-1
7	9	2-3	406	1	US-08-136-993-13
8	9	2-3	410	4	US-10-022-809A-2
9	9	2-3	410	4	US-10-022-809A-5
10	9	2-4	605	4	US-09-976-594-616
11	9	2-4	908	4	US-08-714-741-44
12	8	2-1	81	4	US-09-621-976-5662
					Sequence 7889, Ap
					Sequence 1, Appli
					Sequence 1, Appli
					Sequence 3, Appli
					Sequence 3, Appli
					Sequence 1, Appli
					Sequence 13, Appli
					Sequence 2, Appli
					Sequence 5, Appli
					Sequence 616, App
					Sequence 44, Appl
					Sequence 5662, Ap

13	8	2-1	82	3	US-09-188-930-301	Sequence 301, App
14	8	2-1	82	4	US-09-312-283C-301	Sequence 301, App
15	8	2-1	202	4	US-09-252-991A-29946	Sequence 29946, A
16	8	2-1	249	4	US-09-252-991A-28250	Sequence 28250, A
17	8	2-1	271	4	US-09-489-039A-9994	Sequence 9994, Ap
18	8	2-1	460	4	US-09-252-991A-31090	Sequence 31090, A
19	8	2-1	495	4	US-09-252-991A-24229	Sequence 24229, A
20	8	2-1	606	4	US-08-891-298-3	Sequence 3, Appli
21	8	2-1	606	4	US-08-853-648A-11	Sequence 11, Appli
22	8	2-1	606	4	US-09-564-418-10	Sequence 10, Appli
23	8	2-1	629	4	US-09-134-001C-4394	Sequence 4394, Ap
24	8	2-1	717	4	US-09-644-460-37	Sequence 37, Appli
25	8	2-1	747	3	US-09-035-648-18	Sequence 18, Appli
26	8	2-1	747	3	US-09-001-951-18	Sequence 18, Appli
27	8	2-1	747	4	US-08-818-829-18	Sequence 18, Appli
28	8	2-1	751	4	US-09-252-991A-27424	Sequence 27424, A
29	8	2-1	801	1	US-07-906-349A-6	Sequence 6, Appli
30	8	2-1	808	4	US-09-134-000C-4643	Sequence 4643, Ap
31	8	2-1	1400	4	US-08-630-915A-37	Sequence 37, Appli
32	8	2-1	1497	4	US-09-060-854B-2	Sequence 2, Appli
33	8	2-1	2508	4	US-09-627-650B-7	Sequence 7, Appli
34	8	2-1	2508	4	US-09-436-063C-7	Sequence 7, Appli
35	8	2-1	2544	4	US-09-627-650B-3	Sequence 3, Appli
36	8	2-1	2544	4	US-09-436-063C-3	Sequence 3, Appli
37	8	2-1	2601	4	US-09-627-650B-9	Sequence 9, Appli
38	8	2-1	2601	4	US-09-436-063C-9	Sequence 9, Appli
39	7	1-8	27	1	US-08-425-069-56	Sequence 56, Appli
40	7	1-8	27	2	US-08-317-844B-56	Sequence 56, Appli
41	7	1-8	46	3	US-08-905-223-390	Sequence 390, App
42	7	1-9	54	4	US-09-117-121-30	Sequence 30, Appli
43	7	1-8	65	3	US-09-238-303-15	Sequence 15, Appli
44	7	1-8	65	4	US-09-946-239-15	Sequence 15, Appli
45	7	1-8	68	4	US-09-462-478A-15	Sequence 15, Appli
46	7	1-8	73	4	US-09-134-000C-5438	Sequence 5438, Ap
47	7	1-8	79	4	US-09-621-976-6091	Sequence 6091, Ap
48	7	1-9	83	4	US-09-252-991A-18845	Sequence 18845, A
49	7	1-8	92	4	US-09-343-529-2	Sequence 2, Appli
50	7	1-9	96	4	US-09-489-039A-8081	Sequence 8081, Ap
51	7	1-9	98	2	US-07-814-220-2	Sequence 2, Appli
52	7	1-9	98	2	US-07-812-421-2	Sequence 2, Appli
53	7	1-8	99	3	US-08-905-223-278	Sequence 278, App
54	7	1-9	99	4	US-09-328-352-8191	Sequence 8191, Ap
55	7	1-8	100	4	US-09-621-976-5460	Sequence 5460, Ap
56	7	1-8	108	3	US-09-177-249-172	Sequence 172, App
57	7	1-9	111	3	US-09-034-916-6	Sequence 6, Appli
58	7	1-9	115	4	US-09-543-681A-5637	Sequence 5637, Ap
59	7	1-9	122	4	US-09-540-014-2	Sequence 2, Appli
60	7	1-8	123	4	US-09-252-991A-26903	Sequence 26903, A
61	7	1-8	128	4	US-09-107-532A-4290	Sequence 4290, Ap
62	7	1-9	130	1	US-08-053-006-1	Sequence 1, Appli
63	7	1-8	130	4	US-09-540-014-6	Sequence 6, Appli
64	7	1-8	140	4	US-09-252-991A-30876	Sequence 30876, A
65	7	1-8	146	4	US-09-252-991A-17806	Sequence 17806, A
66	7	1-8	154	4	US-09-489-039A-10927	Sequence 10927, A
67	7	1-8	154	4	US-09-621-976-4212	Sequence 4212, Ap
68	7	1-9	154	4	US-09-621-976-4212	Sequence 4212, Ap
69	7	1-8	156	4	US-09-252-991A-26692	Sequence 26692, A
70	7	1-8	158	4	US-09-252-991A-24956	Sequence 24956, A
71	7	1-8	158	4	US-09-621-976-3906	Sequence 3906, Ap
72	7	1-8	158	4	US-09-621-976-4792	Sequence 4792, Ap
73	7	1-8	160	4	US-09-252-991A-27091	Sequence 27091, A
74	7	1-8	162	2	US-08-319-704-6	Sequence 6, Appli
75	7	1-8	164	4	US-09-252-991A-28523	Sequence 28523, A
76	7	1-9	165	4	US-09-252-991A-28858	Sequence 28858, A
77	7	1-9	172	4	US-09-107-532A-5947	Sequence 5947, Ap
78	7	1-8	176	4	US-09-252-991A-30148	Sequence 30148, A
79	7	1-8	176	4	US-09-198-452A-1032	Sequence 1032, Ap
80	7	1-9	180	4	US-09-252-991A-18500	Sequence 18500, A
81	7	1-8	183	6	5168049-3	Patent No. 5168049
82	7	1-8	186	4	US-09-252-991A-17637	Sequence 17637, A
83	7	1-9	195	4	US-09-071-035-346	Sequence 346, App
84	7	1-8	195	4	US-09-328-352-4743	Sequence 4743, Ap
85	7	1-8	202	4	US-09-134-000C-3660	Sequence 3660, Ap

Mon Sep 20 11:21:27 2004

Sequence 2012, Ap  
Sequence 16621, A  
Sequence 30140, A  
Sequence 23, Appl  
Sequence 20560, A  
Sequence 6557, Ap  
Sequence 18819, A  
Sequence 27694, A  
Sequence 2217, Ap  
Sequence 26289, A  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 8, Appli  
Sequence 8, Appli  
Sequence 8, Appli  
Sequence 19935, A  
Sequence 25654, A  
Sequence 18461, A  
Sequence 10, Appl  
Sequence 1813, Ap  
Sequence 20037, A  
Sequence 1813, Ap  
Sequence 12, Appl  
Sequence 30100, A  
Sequence 18685, A  
Sequence 7987, Ap  
Sequence 6320, Ap  
Sequence 24558, A  
Sequence 4193, Ap  
Sequence 4859, Ap  
Sequence 9102, Ap  
Sequence 4610, Ap  
Sequence 3456, Ap  
Sequence 34, Appl  
Sequence 34, Appl  
Sequence 34, Appl  
Sequence 2871, Ap  
Sequence 6924, Ap  
Sequence 76, Appl  
Sequence 7073, Ap  
Sequence 21, Appl  
Sequence 21, Appl  
Sequence 23, Appl  
Sequence 4223, Ap  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli

4 US-09-540-236-2012  
4 US-09-252-991A-16621  
4 US-09-252-991A-30140  
1 US-08-621-081A-23  
4 US-09-252-991A-20560  
4 US-09-543-681A-6557  
4 US-09-252-991A-18819  
4 US-09-252-991A-27694  
4 US-09-540-236-2217  
4 US-09-252-991A-26289  
2 US-09-151-611-1  
3 US-09-370-102-1  
3 US-08-924-747-8  
3 US-09-247-373B-8  
3 US-09-296-715-8  
4 US-09-252-991A-19935  
4 US-09-252-991A-25654  
4 US-09-252-991A-18461  
3 US-09-234-332-10  
4 US-09-702-705-1813  
4 US-09-736-457-1813  
4 US-09-252-991A-20037  
4 US-09-671-325-1813  
4 US-09-552-225A-12  
4 US-09-252-991A-30100  
4 US-09-252-991A-18685  
4 US-09-489-039A-7987  
4 US-09-134-000C-6320  
4 US-09-252-991A-24558  
4 US-09-134-001C-4193  
4 US-09-328-352-4859  
2 US-08-432-027A-4  
2 US-09-489-039A-9102  
4 US-09-134-000C-4610  
4 US-09-134-001C-3456  
4 US-09-186-276B-34  
4 US-08-842-445-34  
4 US-09-186-188B-34  
4 US-09-134-001C-2871  
4 US-09-543-681A-6924  
4 US-09-152-060-76  
4 US-09-543-681A-7073  
4 US-08-311-731A-21  
4 US-09-725-311-2  
3 US-09-071-224-21  
3 US-09-071-224-21  
3 US-09-543-681A-4223  
3 US-08-413-974-2  
3 US-08-434-418-2  
3 US-08-433-288-2

ALIGNMENTS

RESULT 1  
US-09-328-352-7889  
; Sequence 7889, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GT99-03PA  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 7889  
; LENGTH: 269  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-7889

Alignment Scores:  
QY 349 CCTTACCGTCCAGTATGGAGAGCT 375  
Db 19 ProTyArgProSerTyrglyArgAla 27

Pred. No.: 6.44 Length: 269  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.32% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-328-352-7889 (1-269)  
QY 1050 GTAGCTGTGGCCCTCTCACCAGCGGG 1076  
Db 17 ValAlaValAlaLeuLeuThrSerGly 25

RESULT 2  
US-08-136-993-1  
; Sequence 1, Application US/08136993  
; Patent No. 5430025  
; GENERAL INFORMATION:  
; APPLICANT: Takagi, Hiroshi  
; APPLICANT: Arafuka, Shino  
; APPLICANT: Matsui, Hiroshi  
; APPLICANT: Washizu, Kinya  
; APPLICANT: Ando, Keiichi  
; APPLICANT: Koikeda, Satoshi  
; TITLE OF INVENTION: Recombinant transglutaminase  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue  
; CITY: N.W. Washington, D.C.  
; STATE: Washington, D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20037-3202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/136,993  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/777,447  
; FILING DATE:  
; APPLICATION NUMBER: JP 2-282566  
; FILING DATE: 19-OCT-1990  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-293-7060  
; TELEFAX: 202-293-7860  
; TELEX: 6491103  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 331 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-136-993-1

Alignment Scores:  
Pred. No.: 6.33 Length: 331  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.32% Indels: 0  
DB: 1 Gaps: 0

US-09-729-264-3 (1-1168) x US-08-136-993-1 (1-331)  
QY 349 CCTTACCGTCCAGTATGGAGAGCT 375  
Db 19 ProTyArgProSerTyrglyArgAla 27



```
RESULT 3
US-09-109-063-1
; Sequence 1, Application US/09109063
; Patent No. 6013498
; GENERAL INFORMATION:
; APPLICANT: YOKOYAMA, KEIICHI
; APPLICANT: NAKAMURA, NAAMI
; APPLICANT: MIWA, TETSUYA
; APPLICANT: SEGURO, KATSUYA
; TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
; FILE REFERENCE: 0010-0937-0
; CURRENT APPLICATION NUMBER: US/09/109,063
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: JP 180010/1997
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial
; OTHER INFORMATION: Sequence: TRANSGLUTAMINASE
US-09-109-063-1

Alignment Scores:
Pred. No.: 6.33 Length: 331
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.32% Indels: 0
DB: 3 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-109-063-1 (1-331)
QY 349 CTTACCGTCCAAGTTATGGAGAGCT 375
Db 19 ProTyrArgProSerTyrGlyArgAla 27

RESULT 4
US-08-793-426A-3
; Sequence 3, Application US/08793426A
; Patent No. 6100053
; GENERAL INFORMATION:
; APPLICANT: Bech, Lisbeth
; APPLICANT: No. 6100053revang, Iben
; APPLICANT: Halkier, Torben
; APPLICANT: Rasmussen, Grethe
; APPLICANT: Schafer, Thomas
; APPLICANT: Andersen, Jens
; TITLE OF INVENTION: Microbial Transglutaminases, Their
; TITLE OF INVENTION: Production And Use
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6100053o No. 6100053disk of No. 6100053th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,426A
; FILING DATE: 25-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.
; REGISTRATION NUMBER: 36,993

US-09-729-264-3 (1-1168) x US-08-793-426A-3 (1-331)
QY 349 CTTACCGTCCAAGTTATGGAGAGCT 375
Db 19 ProTyrArgProSerTyrGlyArgAla 27

RESULT 5
US-09-294-565-3
; Sequence 3, Application US/09294565
; Patent No. 6190879
; GENERAL INFORMATION:
; APPLICANT: Bech, Lisbeth
; APPLICANT: No. 6190879revang, Iben
; APPLICANT: Halkier, Torben
; APPLICANT: Rasmussen, Grethe
; APPLICANT: Schafer, Thomas
; APPLICANT: Andersen, Jens
; TITLE OF INVENTION: Microbial Transglutaminases, Their
; TITLE OF INVENTION: Production And Use
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6190879o No. 6190879disk of No. 6190879th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/294,565
; FILING DATE: 19-APR-1999
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4211.224-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

US-09-294-565-3

Alignment Scores: 6.33 Length: 331  
 Pred. No.: 9.00 Matches: 9  
 Score: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: 3 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-294-565-3 (1-331)

QY 349 CCTTACCGTCCAAAGTTATGGAGAGCT 375

Db 19 ProTyrArgProSerTyrGlyArgAla 27

RESULT 6

US-09-448-310-1  
 ; Sequence 1, Application US/09448310  
 ; Patent No. 6538122  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YOKOYAMA, KEIICHI  
 ; APPLICANT: NAKAMURA, NAMI  
 ; APPLICANT: MIWA, TETSUYA  
 ; APPLICANT: SEGURO, KATSUYA  
 ; TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE  
 ; FILE REFERENCE: 0010-0937-0  
 ; CURRENT APPLICATION NUMBER: US/09/448,310  
 ; CURRENT FILING DATE: 1999-11-24  
 ; PRIOR APPLICATION NUMBER: 09/109,063  
 ; PRIOR FILING DATE: 1998-07-02  
 ; NUMBER OF SEQ ID NOS: 62  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 331  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial  
 ; OTHER INFORMATION: Sequence:TRANSGLUTAMINASE  
 US-09-448-310-1

Alignment Scores: 6.33 Length: 331  
 Pred. No.: 9.00 Matches: 9  
 Score: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-448-310-1 (1-331)

QY 349 CCTTACCGTCCAAAGTTATGGAGAGCT 375

Db 19 ProTyrArgProSerTyrGlyArgAla 27

RESULT 7

US-08-136-993-13  
 ; Sequence 13, Application US/08136993  
 ; Patent No. 5420025  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Takagi, Hiroshi  
 ; APPLICANT: Arafuka, Shino  
 ; APPLICANT: Matsui, Hiroshi  
 ; APPLICANT: Washizu, Kinya  
 ; APPLICANT: Ando, Keiichi  
 ; APPLICANT: Koikeda, Satoshi  
 ; TITLE OF INVENTION: Recombinant transglutaminase  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
 ; STREET: 2100 Pennsylvania Avenue  
 ; CITY: N.W.

STATE: Washington, D.C.  
 COUNTRY: U.S.A.  
 ZIP: 20037-3202  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.24  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/136,993  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/777,447  
 FILING DATE:  
 APPLICATION NUMBER: JP 2-282566  
 FILING DATE: 19-OCT-1990  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-293-7060  
 TELEFAX: 202-293-7860  
 TELEX: 6491103  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 406 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-136-993-13

Alignment Scores: 6.22 Length: 406  
 Pred. No.: 9.00 Matches: 9  
 Score: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: 1 Gaps: 0

US-09-729-264-3 (1-1168) x US-08-136-993-13 (1-406)

QY 349 CCTTACCGTCCAAAGTTATGGAGAGCT 375

Db 94 ProTyrArgProSerTyrGlyArgAla 102

RESULT 8

US-10-022-809A-2  
 ; Sequence 2, Application US/10022809A  
 ; Patent No. 6660510  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LIN, Yi-Shin  
 ; APPLICANT: LIU, Chang-Hsieh  
 ; APPLICANT: CHU, Wen-Shen  
 ; TITLE OF INVENTION: TRANSGLUTAMINASE GENE OF STREPTOVERTICILLUM LADAKANUM AND THE  
 ; FILE REFERENCE: U 013779-2  
 ; CURRENT APPLICATION NUMBER: US/10/022,809A  
 ; CURRENT FILING DATE: 2001-12-17  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 410  
 ; TYPE: PRT  
 ; ORGANISM: Streptovorticillum ladakanum  
 US-10-022-809A-2

Alignment Scores: 6.21 Length: 410  
 Pred. No.: 9.00 Matches: 9  
 Score: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-10-022-809A-2 (1-410)

QY 349 CCTTACGTCCTCAAGTTATGGAGACT 375  
 Db 98 ProTyrArgProSerTyrGlyArgAla 106

RESULT 9  
 US-10-022-809A-5  
 ; Sequence 5, Application US/10022809A  
 ; Patent No. 6660510  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LIN, Yi-Shin  
 ; APPLICANT: LIU, Chang-Hsielsh  
 ; APPLICANT: CHU, Wen-Shen  
 ; TITLE OF INVENTION: TRANSGLUTAMINASE GENE OF STREPTOVERTICILLIUM LADAKANUM AND THE  
 ; FILE REFERENCE: U 013779-2  
 ; CURRENT APPLICATION NUMBER: US/10/022,809A  
 ; CURRENT FILING DATE: 2001-12-17  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 5  
 ; LENGTH: 410  
 ; TYPE: PRT  
 ; ORGANISM: Streptovorticillium ladakanum  
 US-10-022-809A-5

Alignment Scores:  
 Pred. No.: 6.21 Length: 410  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.32% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-10-022-809A-5 (1-410)

QY 349 CCTTACGTCCTCAAGTTATGGAGACT 375  
 Db 98 ProTyrArgProSerTyrGlyArgAla 106

RESULT 10  
 US-09-976-594-616  
 ; Sequence 616, Application US/09976594  
 ; Patent No. 6673549  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Furness, Michael  
 ; APPLICANT: Buchbinder, Jenny  
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
 ; FILE REFERENCE: PA-0041 US  
 ; CURRENT APPLICATION NUMBER: US/09/976,594  
 ; CURRENT FILING DATE: 2001-10-12  
 ; PRIOR APPLICATION NUMBER: 60/240,409  
 ; PRIOR FILING DATE: 2000-10-12  
 ; NUMBER OF SEQ ID NOS: 1143  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 616  
 ; LENGTH: 605  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: Incyte ID No. 6673549 1692213CB1  
 ; NAME/KEY: unsure  
 ; LOCATION: 596  
 ; OTHER INFORMATION: unknown or other  
 US-09-976-594-616

Alignment Scores:  
 Pred. No.: 6 Length: 605  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.38% Indels: 0

DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-976-594-616 (1-605)

QY 829 ACAACACGCGCGCGCAGCAGCA 803  
 Db 523 ThrThrThrAlaAlaAlaAlaAla 531

RESULT 11  
 US-08-714-741-44  
 ; Sequence 44, Application US/08714741  
 ; Patent No. 6500613  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Briles, David E.  
 ; APPLICANT: McDaniel, Larry S.  
 ; APPLICANT: Swiatlo, Edwin  
 ; APPLICANT: Yother, Janet  
 ; APPLICANT: Crain, Marilyn J.  
 ; APPLICANT: Hollingshead, Susan  
 ; APPLICANT: Tart, Rebecca  
 ; APPLICANT: Brooks-Walter, Alexis  
 ; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,  
 ; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,  
 ; TITLE OF INVENTION: PORTIONS AND PRODUCTS  
 ; NUMBER OF SEQUENCES: 47  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Curtis, Morris & Safford, P.C.  
 ; STREET: 530 Fifth Avenue  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/714,741  
 ; FILING DATE: 16-SEP-1996  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Frommer Esq., William S.  
 ; REGISTRATION NUMBER: 25,506  
 ; REFERENCE/DOCKET NUMBER: 454312-2460  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 840-3333  
 ; TELEFAX: (212) 840-0712  
 ; INFORMATION FOR SEQ ID NO: 44:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 908 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 US-08-714-741-44

Alignment Scores:  
 Pred. No.: 5.79 Length: 908  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2.38% Indels: 0  
 DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-08-714-741-44 (1-908)

QY 826 ACAACGCGCGCGCAGCAGCG 800  
 Db 321 ThrThrThrAlaAlaAlaAlaAla 329

RESULT 12  
 US-09-621-976-5662

Mon Sep 20 11:21:27 2004

us-09-729-264-3.oli.ra1

```

; Sequence 5662, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 5662
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -24..-1
US-09-621-976-5662

Alignment Scores:
Pred. No.: 60.4 Length: 81
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.06% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-621-976-5662 (1-81)

QY 820 TCGTGTGGTGGCTGCACTGCTG 843
Db 12 SerLeuLeuPrpLeuGlnLeuLeu 19

RESULT 13
US-09-188-930-301
; Sequence 301, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 301
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Mouse
US-09-188-930-301

Alignment Scores:
Pred. No.: 60.3 Length: 82
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.06% Indels: 0
DB: 3 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-188-930-301 (1-82)

QY 841 CTGCTGCCGTGTGTTTCTGCTG 864
Db 35 LeuLeuProLeuLeuPheLeuLeu 42

RESULT 14

```

```

US-09-312-283C-301
; Sequence 301, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated from Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 301
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Mouse
US-09-312-283C-301

Alignment Scores:
Pred. No.: 60.3 Length: 82
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.06% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-312-283C-301 (1-82)

QY 841 CTGCTGCCGTGTGTTTCTGCTG 864
Db 35 LeuLeuProLeuLeuPheLeuLeu 42

RESULT 15
US-09-252-991A-29946
; Sequence 29946, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29946
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29946

Alignment Scores:
Pred. No.: 55.7 Length: 202
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.12% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-252-991A-29946 (1-202)

QY 146 GTTCCAGCCCTGGGAGACGGTGC 123
Db 9 AlaSerSerProGlyArgArgCys 16

```

## RESULT 16

US-09-252-991A-28250  
; Sequence 28250, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 28250  
; LENGTH: 249  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-28250

Alignment Scores:  
Pred. No.: 54.7 Length: 249  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.12% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-252-991A-28250 (1-249)

QY 146 GCTTCAGCCCTGGAGACGGTGC 123

Db 9 AlaSerSerProGlyArgArgCys 16  
|||||

## RESULT 17

US-09-489-039A-9994  
; Sequence 9994, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 9994  
; LENGTH: 271  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9994

Alignment Scores:  
Pred. No.: 54.3 Length: 271  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.12% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-489-039A-9994 (1-271)

QY 622 CTTTCGGGCTTCAGGCTCTCCA 599

Db 125 LeuAlaGlyLeuGlnAlaLeuPro 132  
|||||

## RESULT 18

US-09-252-991A-33090  
; Sequence 33090, Application US/09252991A

; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 33090  
; LENGTH: 460  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-33090

Alignment Scores:  
Pred. No.: 51.8 Length: 460  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-252-991A-33090 (1-460)

QY 300 GGACATCAGATCGAGCTCCAGA 323

Db 244 GlyThrSerAspAlaAlaSerArg 251  
|||||

## RESULT 19

US-09-252-991A-24229  
; Sequence 24229, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24229  
; LENGTH: 495  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24229

Alignment Scores:  
Pred. No.: 51.5 Length: 495  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.06% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-252-991A-24229 (1-495)

QY 1076 GCTGATCAAGTCCACCCAGGCCA 1099

Db 279 AlaAspGlnArgProArgPro 286  
|||||

## RESULT 20

US-08-891-298-3  
; Sequence 3, Application US/08891298  
; Patent No. 6300488

```

; GENERAL INFORMATION:
; APPLICANT: Gage, Frederick H.
; APPLICANT: Suhr, Steven T.
; TITLE OF INVENTION: Modified Lepidopteran Receptors
; TITLE OF INVENTION: and Hybrid Multi-Functional Proteins for Use in Transcription
; TITLE OF INVENTION: and Transgene Expression Regulation
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,298
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-677-1409
; TELEFAX: 619-677-1465
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 606 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-891-298-3

Alignment Scores:
Pred. No.: 50.6 Length: 606
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.12% Indels: 0
Gaps: 4

US-09-729-264-3 (1-1168) x US-08-891-298-3 (1-606)
QY 846 CAGCAGCAGTTGCAGCCACAA 823
Db 135 GlnGlnLeuGlnProGlnGln 142

RESULT 21
US-08-653-648A-11
; Sequence 11, Application US/08653648A
; Patent No. 6379945
; GENERAL INFORMATION:
; APPLICANT: Jepsen, Ian
; APPLICANT: Greenland, Andrew
; APPLICANT: Martinez, Alberto
; TITLE OF INVENTION: A Gene Switch
; FILE REFERENCE: PPD50047/US
; CURRENT APPLICATION NUMBER: US/08/653,648A
; CURRENT FILING DATE: 1996-05-24
; PRIOR APPLICATION NUMBER: GB 9510759.5
; PRIOR FILING DATE: 1995-05-26
; PRIOR APPLICATION NUMBER: GB 9605656.9

```

```

; PRIOR FILING DATE: 1996-03-18
; PRIOR APPLICATION NUMBER: GB 9513882.2
; PRIOR FILING DATE: 1995-07-07
; PRIOR APPLICATION NUMBER: GB 9517316.7
; PRIOR FILING DATE: 1995-08-24
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 606
; TYPE: PRT
; ORGANISM: Bombyx mori
; US-08-653-648A-11

Alignment Scores:
Pred. No.: 50.6 Length: 606
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.12% Indels: 0
Gaps: 4

US-09-729-264-3 (1-1168) x US-08-653-648A-11 (1-606)
QY 846 CAGCAGCAGTTGCAGCCACAA 823
Db 135 GlnGlnLeuGlnProGlnGln 142

RESULT 22
US-09-564-418-10
; Sequence 10, Application US/09564418
; Patent No. 6610828
; GENERAL INFORMATION:
; APPLICANT: Jepsen, Ian
; APPLICANT: Martinez, Alberto
; APPLICANT: Greenland, Andrew James
; TITLE OF INVENTION: A GENE SWITCH
; FILE REFERENCE: 1392/4/3
; CURRENT APPLICATION NUMBER: US/09/564,418
; CURRENT FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: US 09/564,418
; PRIOR FILING DATE: 2000-05-03
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 606
; TYPE: PRT
; ORGANISM: Bombyx mori
; US-09-564-418-10

Alignment Scores:
Pred. No.: 50.6 Length: 606
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.12% Indels: 0
Gaps: 4

US-09-729-264-3 (1-1168) x US-09-564-418-10 (1-606)
QY 846 CAGCAGCAGTTGCAGCCACAA 823
Db 135 GlnGlnLeuGlnProGlnGln 142

RESULT 23
US-09-134-001C-4394
; Sequence 4394, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007

```

```
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4394
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4394

Alignment Scores:
Pred. No.: 50.4 Length: 629
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.12% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-134-001C-4394 (1-629)
Qy 1012 GCGAGGAGAGAGCGGTCTCTGT 989
Db 444 GlyArgGluArgSerGlyValCys 451

RESULT 24
US-09-644-460-37
; Sequence 37, Application US/09644460
; Patent No. 6657053
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; TITLE OF INVENTION: Reciprocal Subtraction Differential
; FILE REFERENCE: Display
; CURRENT APPLICATION NUMBER: US/09/644,460
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US99/04323
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: US 09/197,889
; PRIOR FILING DATE: 1998-11-23
; PRIOR APPLICATION NUMBER: US 09/185,115
; PRIOR FILING DATE: 1998-11-03
; PRIOR APPLICATION NUMBER: US 09/032,684
; PRIOR FILING DATE: 1998-02-27
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 717
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-644-460-37

Alignment Scores:
Pred. No.: 49.9 Length: 717
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.12% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-644-460-37 (1-717)
Qy 823 ACGACGGCGGACGACGACGACGCG 800
Db 482 ThrThrAlaAlaAlaAlaAlaAla 489

RESULT 25
US-09-035-648-18
; Sequence 18, Application US/09035648
; Patent No. 6100031
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL
; TITLE OF INVENTION: GROWTH AND PROLIFERATION
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/035,648
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/818,829
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-035-648-18

Alignment Scores:
Pred. No.: 49.7 Length: 747
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.12% Indels: 0
DB: 3 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-035-648-18 (1-747)
Qy 1060 GCGACGACTGTTTCTTTGTTC 1037
Db 312 AlaThrAlaThrValSerLeuPhe 319

RESULT 26
US-09-001-951-18
; Sequence 18, Application US/09001951
; Patent No. 6268470
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL
; TITLE OF INVENTION: GROWTH AND PROLIFERATION
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/001,951
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/818,829
; FILING DATE: 14-MAR-1997
; APPLICATION NUMBER: 60/013,438
; FILING DATE: 15-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/003001
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-09-001-951-18
;
; Alignment Scores:
; Pred. No.: 49.7 Length: 747
; Score: 8.00 Matches: 8
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 2.12% Indels: 0
; DB: 3 Gaps: 0
;
; US-09-729-264-3 (1-1168) x US-09-001-951-18 (1-747)
;
; QY 1060 GCCACAGCTACTGTTTCTTTGTTTC 1037
; DB 312 AlaThrAlaThrValSerLeuPhe 319
;
; RESULT 27
; US-08-818-829-18
; Sequence 18, Application US/08818829
; Patent No. 6458939
; GENERAL INFORMATION:
; APPLICANT: Shvian, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,829
; FILING DATE: 14-MAR-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/013,438
; FILING DATE: 15-MAR-1996

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-818-829-18
;
; Alignment Scores:
; Pred. No.: 49.7 Length: 747
; Score: 8.00 Matches: 8
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 2.12% Indels: 0
; DB: 4 Gaps: 0
;
; US-09-729-264-3 (1-1168) x US-08-818-829-18 (1-747)
;
; QY 1060 GCCACAGCTACTGTTTCTTTGTTTC 1037
; DB 312 AlaThrAlaThrValSerLeuPhe 319
;
; RESULT 28
; US-09-252-991A-27424
; Sequence 27424, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27424
; LENGTH: 751
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-27424
;
; Alignment Scores:
; Pred. No.: 49.6 Length: 751
; Score: 8.00 Matches: 8
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 2.06% Indels: 0
; DB: 4 Gaps: 0
;
; US-09-729-264-3 (1-1168) x US-09-252-991A-27424 (1-751)
;
; QY 701 AGTTTACCGAGTTTAGGTTTTC 724
; DB 560 SerLeuProSerLeuGlyPheSer 567
;
; RESULT 29
; US-07-906-349A-6
; Sequence 6, Application US/07906349A
; Patent No. 5434064
; GENERAL INFORMATION:
; APPLICANT: Schlessinger, Joseph

```



```
/ APPLICANT: Skolnik, Edward Y.
/ APPLICANT: Margolis, Benjamin L.
/ TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR
/ TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE KINASES AN
/ TITLE OF INVENTION: TARGET PROTEINS
/ NUMBER OF SEQUENCES: 16
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Browdy and Neimark
/ STREET: 419 Seventh Street, N.W.
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: USA
/ ZIP: 20004
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/906,349A
/ FILING DATE: 30-JUN-1992
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/643,237
/ FILING DATE: 18-JAN-1991
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 202-628-5197
/ TELEFAX: 202-737-3528
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 801 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-07-906-349A-6

Alignment Scores:
Pred. No.: 49.4 Length: 801
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.12% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-3 (1-1168) x US-07-906-349A-6 (1-801)
Qy 826 ACAACGAGCGGCGAGCAGCAGCA 803
Db 464 ThrThrAlaAlaAlaAlaAla 471

RESULT 30
US-09-134-000C-4643
/ Sequence 4643, Application US/09134000C
/ Patent No. 6617156
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
/ TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 032796-032
/ CURRENT APPLICATION NUMBER: US/09/134,000C
/ CURRENT FILING DATE: 1998-08-13
/ PRIOR APPLICATION NUMBER: US 60/055,778
/ PRIOR FILING DATE: 1997-08-15
/ NUMBER OF SEQ ID NOS: 6812
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 4643
/ LENGTH: 808
/ TYPE: PRT
/ ORGANISM: Enterococcus faecalis
US-09-134-000C-4643

Alignment Scores:
Pred. No.: 49.4 Length: 808
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.12% Indels: 0
DB: 1 Gaps: 0

US-09-729-264-3 (1-1168) x US-07-906-349A-6 (1-801)
Qy 826 ACAACGAGCGGCGAGCAGCAGCA 803
Db 464 ThrThrAlaAlaAlaAlaAla 471

RESULT 31
US-08-630-915A-37
/ Sequence 37, Application US/08630915A
/ Patent No. 6309820
/ GENERAL INFORMATION:
/ APPLICANT: SPARKS, Andrew B.
/ APPLICANT: HOFFMAN, No. 6309820h
/ APPLICANT: KAY, Brian K.
/ APPLICANT: FOWLES, Dana M.
/ APPLICANT: MCCONNELL, Stephen J.
/ TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
/ TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
/ NUMBER OF SEQUENCES: 227
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Pennie & Edmonds LLP
/ STREET: 1155 Avenue of the Americas
/ CITY: New York
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10036-2711
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/630,915A
/ FILING DATE: 03-APR-1996
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Mistrock, S. Leslie
/ REGISTRATION NUMBER: 18,872
/ REFERENCE/DOCKET NUMBER: 1101-174
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 790-9090
/ TELEFAX: (212) 869-8864/9741
/ TELEX: 66141 PENNIE
/ INFORMATION FOR SEQ ID NO: 37:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1400 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA
US-08-630-915A-37

Alignment Scores:
Pred. No.: 47 Length: 1400
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.12% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-08-630-915A-37 (1-1400)
Qy 835 GCAGCCACAAACGACGCGCA 812
Db 737 AlaAlaThrThrThrAlaAla 744
```

RESULT 32  
US-09-060-854B-2  
; Sequence 2, Application US/09060854B  
; Patent No. 6642011  
; GENERAL INFORMATION:  
; APPLICANT: Estell, David Aaron  
; TITLE OF INVENTION: Human Protease and Use of Such Protease for Pharmaceutical  
; TITLE OF INVENTION: Applications and for Reducing the Allergenicity of No. 6642011-H  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: GCS32  
; CURRENT APPLICATION NUMBER: US/09/060,854B  
; PRIOR FILING DATE: 1998-04-15  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1497  
; TYPE: PRT  
; ORGANISM: B. amyloliquefaciens  
US-09-060-854B-2

Alignment Scores:  
Pred. No.: 46.7 Length: 1497  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.12% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-060-854B-2 (1-1497)

QY 835 GCAGCCACACACGCGCGGCA 812  
Db 1342 ALAAlaThrThrThrAlaAla 1349

RESULT 33  
US-09-627-650B-7  
; Sequence 7, Application US/09627650B  
; Patent No. 6406872  
; GENERAL INFORMATION:  
; APPLICANT: Bamber, Bruce  
; APPLICANT: Jorgensen, Erik  
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and  
; TITLE OF INVENTION: Methods Related Thereto  
; FILE REFERENCE: 21101.0009U3  
; CURRENT APPLICATION NUMBER: US/09/627,650B  
; CURRENT FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 09/436,063  
; PRIOR FILING DATE: 1999-11-08  
; PRIOR APPLICATION NUMBER: 60/107,727  
; PRIOR FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 2508  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-09-627-650B-7

Alignment Scores:  
Pred. No.: 44.7 Length: 2508  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.12% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-627-650B-7 (1-2508)

QY 823 ACGACGCGGCAGCAGCAGCG 800  
Db 2491 ThrThrAlaAlaAlaAlaAla 2498

RESULT 34  
US-09-436-063C-7  
; Sequence 7, Application US/09436063C  
; Patent No. 6407210  
; GENERAL INFORMATION:  
; APPLICANT: Bamber, Bruce  
; APPLICANT: Jorgensen, Erik  
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and  
; TITLE OF INVENTION: Methods Related Thereto  
; FILE REFERENCE: P-1095corrected  
; CURRENT APPLICATION NUMBER: US/09/436,063C  
; CURRENT FILING DATE: 1999-11-08  
; PRIOR APPLICATION NUMBER: 60/107727  
; PRIOR FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 2508  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-09-436-063C-7

Alignment Scores:  
Pred. No.: 44.7 Length: 2508  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.12% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-436-063C-7 (1-2508)

QY 823 ACGACGCGGCAGCAGCAGCG 800  
Db 2491 ThrThrAlaAlaAlaAlaAla 2498

RESULT 35  
US-09-627-650B-3  
; Sequence 3, Application US/09627650B  
; Patent No. 6406872  
; GENERAL INFORMATION:  
; APPLICANT: Bamber, Bruce  
; APPLICANT: Jorgensen, Erik  
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and  
; TITLE OF INVENTION: Methods Related Thereto  
; FILE REFERENCE: 21101.0009U3  
; CURRENT APPLICATION NUMBER: US/09/627,650B  
; CURRENT FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 09/436,063  
; PRIOR FILING DATE: 1999-11-08  
; PRIOR APPLICATION NUMBER: 60/107,727  
; PRIOR FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 2544  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-09-627-650B-3

Alignment Scores:  
Pred. No.: 44.6 Length: 2544  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.12% Indels: 0  
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-627-650B-3 (1-2544)

QY 823 ACGACGCGGCAGCAGCAGCG 800  
Db 2527 ThrThrAlaAlaAlaAlaAla 2534

```
RESULT 36
US-09-436-063C-3
; Sequence 3, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436,063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2544
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-3

Alignment Scores:
Pred. No.: 44.6 Length: 2544
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.12% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-436-063C-3 (1-2544)
QY 823 ACGACGGCGGCGACGACGACGCG 800
Db 2527 ThrThrAlaAlaAlaAlaAla 2534

RESULT 37
US-09-627-650B-9
; Sequence 9, Application US/09627650B
; Patent No. 6406872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2601
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-9

Alignment Scores:
Pred. No.: 44.5 Length: 2601
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.12% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-627-650B-9 (1-2601)
QY 823 ACGACGGCGGCGACGACGACGCG 800
|||||
```

```
Db 2584 ThrThrAlaAlaAlaAlaAla 2591

RESULT 38
US-09-436-063C-9
; Sequence 9, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; FILE REFERENCE: P-1095corrected
; CURRENT APPLICATION NUMBER: US/09/436,063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2601
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-9

Alignment Scores:
Pred. No.: 44.5 Length: 2601
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.12% Indels: 0
DB: 4 Gaps: 0

US-09-729-264-3 (1-1168) x US-09-436-063C-9 (1-2601)
QY 823 ACGACGGCGGCGACGACGACGCG 800
|||||
Db 2584 ThrThrAlaAlaAlaAlaAla 2591

RESULT 39
US-08-425-069-56
; Sequence 56, Application US/08425069
; Patent No. 5728810
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Hinman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5728810th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/425,069
; APPLICATION NUMBER: US/08/425,069
; FILING DATE: 19-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-106P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
```



OTHER INFORMATION: seq IKCSSWISSLASG/IP  
US-08-905-223-390

Alignment Scores: Length: 46  
Pred. No.: 535 Matches: 7  
Score: 7.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 1.80%  
DB: 3 Indels: 0  
Gaps: 0

US-09-729-264-3 (1-1168) x US-08-905-223-390 (1-46)

QY 710 AGTTAGTTTTCATTGCT 730

Db 38 SerLeuGlyPheSerLeuPro 44

RESULT 42

US-09-117-121-30  
; Sequence 30, Application US/09117121  
; Patent No. 6307020  
; GENERAL INFORMATION:  
; APPLICANT: Hew, Choy  
; APPLICANT: Gong, Zhiyuan  
; TITLE OF INVENTION: Intracellular Antifreeze Polypeptides  
; TITLE OF INVENTION: and Nucleic Acids  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/117,121  
; FILING DATE: 20-NOV-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/CA97/00062  
; FILING DATE: 30-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Kenneth A.  
; REGISTRATION NUMBER: 31,677  
; REFERENCE/DOCKET NUMBER: 016252-001610US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 54 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-09-117-121-30

Alignment Scores: Length: 54  
Pred. No.: 528 Matches: 7  
Score: 7.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 1.85%  
DB: 4 Indels: 0  
Gaps: 0

US-09-729-264-3 (1-1168) x US-09-117-121-30 (1-54)

QY 820 ACSCGCGCAGCAGCAGCG 800

|||||

Db 22 ThrAlaAlaAlaAlaAla 28

RESULT 43

US-09-238-303-15  
; Sequence 15, Application US/09238303B  
; Patent No. 6284253  
; GENERAL INFORMATION:  
; APPLICANT: Barr, Margaret C.  
; TITLE OF INVENTION: No. 6284253el Feline Immunodeficiency Virus Nucleotide Sequence  
; FILE REFERENCE: 18617.0059  
; CURRENT APPLICATION NUMBER: US/09/238,303B  
; EARLIER FILING DATE: 1999-01-28  
; EARLIER APPLICATION NUMBER: US 60/072,927  
; EARLIER FILING DATE: 1998-01-29  
; NUMBER OF SEQ ID NOS: 17  
; SEQ ID NO 15  
; LENGTH: 65  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: protein encoded by the open reading frame (orfe) of a  
; OTHER INFORMATION: recombinant viral clone constructed from the genomic DNA of a  
; OTHER INFORMATION: Pallas's cat feline immunodeficiency virus  
US-09-238-303-15

Alignment Scores:

Pred. No.: 519 Length: 65  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 1.80%  
DB: 3 Indels: 0  
Gaps: 0

US-09-729-264-3 (1-1168) x US-09-238-303-15 (1-65)

QY 866 AGAAGAAAAGAGGATTCGT 886

Db 24 ArgArgLysArgGlyPheArg 30

RESULT 44

US-09-946-239-15  
; Sequence 15, Application US/09946239  
; Patent No. 6579527  
; GENERAL INFORMATION:  
; APPLICANT: Barr, Margaret C.  
; TITLE OF INVENTION: No. 6579527el Feline Immunodeficiency Virus Nucleotide and  
; TITLE OF INVENTION: Polypeptide Sequences  
; FILE REFERENCE: 18617.0059  
; CURRENT APPLICATION NUMBER: US/09/946,239  
; CURRENT FILING DATE: 2001-09-04  
; PRIOR APPLICATION NUMBER: US 09/238,303, US 60/072,927  
; PRIOR FILING DATE: 1999-01-28, 1998-01-29  
; NUMBER OF SEQ ID NOS: 17  
; SEQ ID NO 15  
; LENGTH: 65  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: protein encoded by the open reading frame (orfe) of a  
; OTHER INFORMATION: recombinant viral clone constructed from the genomic DNA of a  
; OTHER INFORMATION: Pallas's cat feline immunodeficiency virus  
US-09-946-239-15

Alignment Scores:

Pred. No.: 519 Length: 65  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 1.80%  
DB: 4 Indels: 0  
Gaps: 0

US-09-729-264-3 (1-1168) x US-09-946-239-15 (1-65)

```
QY      866 AGAAGRAAAGAGGATTTCGT 886
Db      24 ArgArgLysArgGlyPheArg 30

RESULT 45
US-09-462-478A-15
; Sequence 15, Application US/09462478A
; Patent No. 6541607
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND
; TITLE OF INVENTION:  SUBLANCIN LANTIBIOTIC PRODUCED BY BACILLUS SUBTILIS 168
; FILE REFERENCE: 8172-8072
; CURRENT APPLICATION NUMBER: US/09/462,478A
; CURRENT FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: US/60/053,035
; PRIOR FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Cytolysin L1
US-09-462-478A-15

Alignment Scores:
Pred. No.:      517      Length:      68
Score:          7.00     Matches:      7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:      1.80% Indels:      0
DB:              4      Gaps:      0

US-09-729-264-3 (1-1168) x US-09-462-478A-15 (1-68)

QY      828 GTGGCTGCAACTGCTGCGCC 848
Db      37 ValAlaAlaThrAlaAlaAla 43

Search completed: September 18, 2004, 23:05:12
Job time : 34.8543 secs
```